The sperm proteome of the red abalone, *Haliotis rufescens,* and the discovery and characterization of an abundant, rapidly evolving acrosome protein

Melody Rynerson Palmer

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Willie J. Swanson, Chair
Michael J. MacCoss
Maitreya J. Dunham

Program Authorized to Offer Degree:
Genome Sciences

University of Washington

Abstract

The sperm proteome of the red abalone, *Haliotis rufescens,* and the discovery and characterization of an abundant, rapidly evolving acrosome protein

Melody Rynerson Palmer

Chair of Supervisory Committee:
Professor Willie J. Swanson
Department of Genome Sciences

Abalone, a broadcast spawning marine mollusk, is an important model for molecular interactions and positive selection in fertilization, but the focus has previously been on only two sperm proteins, lysin and sp18. We have used genomic and proteomic techniques to bring new insights to this model by characterizing the testis transcriptome and sperm proteome of the red abalone *Haliotis rufescens*. One pair of homologous testis-specific proteins contains a secretion signal and is small, abundant, and associated with the acrosome, here named sp6_4D and sp6_8D. Comparative analysis revealed that homologs are extremely divergent between species, and show strong evidence for positive selection. The acrosomal localization and rapid evolution of these proteins indicates that they play an important role in fertilization and could be involved in the species-specificity of sperm-egg interaction in abalone. Our genomic and proteomic characterization of abalone fertilization resulted in the identification of these interesting novel peptides that have eluded detection in this important model system for twenty years. Through additional work on these proteins, we have created a foundation for the future discovery of their interacting proteins and further investigations into their function.

Table of Contents

Acknowledgements

I could not have completed the research for this dissertation without the help of many people. My advisor, Willie Swanson, provided continuous advice, support, and encouragement. Members of the Swanson Lab participated in helpful discussions, providing advice, perspectives, and ideas: Jan Aagaard, Katrina Claw, Geoff Findlay, Joe Gasper, Renee George, Jennifer McCreight, and Stevan Springer. In addition, I would like to acknowledge the members of my committee for their comments and productive discussions throughout my time in graduate school: Maitreya Dunham, Lorenz Hauser, Michael MacCoss, and Barbara Wakimoto.

The following people contributed significantly to particular aspects of the project:
Mass spectrometry: Genn Merrihew from the MacCoss Lab provided training, advice for sample preparation, and ran the instruments.
RNA-seq: From the Shendure Lab, Joe Hiatt helped with sample preparation, and Choli Lee operated the sequencer. Renee George provided help with data analysis.
I mentored two excellent undergraduate researchers, whose work contributed to this project: Aleena Ouaddi and Lia Stewart.

I would also like to thank my family and friends for their personal support, especially my parents, Kay Rynerson-McCabe and David Rynerson, and my husband Ira.

Chapter 1

Evolution of sperm-egg interaction

A version of this chapter was previously published in the book
'Evolution in the fast lane: rapidly evolving genes and genetic systems,' ©2012 Oxford University Press, Edited by Rama S. Singh, Jianping Xu, and Rob J. Kulathinal.

1.1 Introduction

The success of fertilization and reproduction in sexual species is dependent on the interaction between sperm and egg. Some sperm-egg interactions succeed between distant species, creating hybrids, but others fail between closely related species at one or more stages. Given the importance of fertilization in the maintenance of species integrity and boundaries, it is surprising that reproductive proteins are often evolving more rapidly than the genome-wide average (Swanson and Vacquier 2002). In this review we will discuss the evolution of sperm-egg interaction in terms of the mechanisms, proteins, and selective pressures involved. Experimental approaches are rapidly changing with new genomic and proteomic technologies, and there is a correspondingly rapid increase in the amount of new data on fertilization-associated proteins. This data will improve our understanding of the evolution of sperm-egg interacting proteins, for which variation or conservation may determine the degree of reproductive isolation between populations or species. There are exciting opportunities to identify the molecular players in the process of speciation: the more species studied in greater detail, the more proteins could be identified.

2.2 Evolution at each step of sperm-egg interaction

Sperm and eggs in the organisms we will discuss are similar in their basic structures, but diverse in their molecules and mechanisms of recognition, binding, entry, and fusion. At each step, we see variation in the degree of species specificity and rate of protein evolution. Is this a result of variation in selective pressure at the different stages of fertilization? Looking at the specific steps (Figure 1.1) will give insights into the stages of fertilization that are under the strongest selection in rapidly evolving systems. For example, the marine mollusk abalone sperm proteins lysin and sp18 are both evolving rapidly, but sp18, which acts closer to the end of fertilization, has higher divergence between species than lysin (Metz et al. 1998). Non-sperm ejaculated proteins, such as seminal fluid proteins (sfps) in insects (*Drosophila* - (Findlay et al. 2009)*)* and mammals (rodents - (Ramm et al. 2009)) also evolve rapidly, and their function is the subject of active research.

In order to physically interact, the sperm must first find the egg. An important factor in this step is chemoattraction – where the egg releases an attractant that stimulates sperm to swim in the direction of the egg. Chemotaxis of sperm to eggs has been observed *in vitro* in organisms from free-spawning marine invertebrates to mammals (Eisenbach 1999). A diverse range of *in vitro* attractants has been identified, including small molecules, peptides, low molecular mass proteins, and a lipid molecule, but few have been associated with eggs outside of marine invertebrates (Ward et al. 1985; Miller and Vogt 1996; Spehr et al. 2003; Kaupp et al. 2008). The sperm's environment and path to the egg differ between internal and external fertilizers, which may impact the evolution of the attractant molecules and receptor proteins. In internal fertilizers such as mammals, *in vivo* chemoattraction is not well understood. Researchers have identified a few

candidate receptors in mammalian sperm, but none are confirmed or well characterized. One group of candidates is testis-expressed olfactory receptors. A small group of these are implicated in sperm signal response, such as hOR17-4, but no natural ligands are known (Spehr et al. 2003). In mammals, testis-expressed ORs are highly conserved, particularly in the ligand binding domain (Branscomb et al. 2000). Consistent with this, Sun et al. 2003 showed that chemotaxis toward follicular factors is not species specific in rabbit and human sperm (Sun et al. 2003).

Chemoattraction is particularly important in free-spawning marine organisms because of the heightened possibility of encountering heterospecific gametes and the dilution of sperm and eggs in the ocean. For example, corals undergo mass spawning events where gametes from multiple species are present in the water at once. One study showed that three species of *Acropora* have species-specific sperm motility initiation by eggs, but no molecular players are know n(Morita et al. 2006). In abalone, chemotaxis is also species specific (Riffell et al. 2004). *Haliotis rufescens* eggs release L-tryptophan, which acts as a chemoattractant. Sperm from *H. fulgens* do not respond to egg factors from *H. rufescens*, and the same for the reciprocal. This could be a result of selection against hybridization in species with overlapping habitats and spawning seasons. In sea urchin species, the egg peptide speract varies in sequence and sperm response (Guerrero et al. 2010). Conversely, in frogs and mice, allurin, a member of the Crisp protein family, appears to have conserved chemotactic induction (Burnett et al. 2008). More candidate receptor proteins will be discovered with new proteomic techniques. More sequence information may reveal interesting patterns of selection at the stage of chemotaxis.

The next step is binding and passage through the egg envelope, one of the best-

studied molecular mechanisms of sperm-egg interaction. The known molecules that mediate this interaction are diverse. Egg envelopes are composed of large glycoproteinaceous matrices. They share structural properties across taxa from marine invertebrates to mammals, such as the ZP domain (Monné et al. 2006), but are divergent in sequence. A recent analysis showed similarity in the functionally important residues of structurally homologous regions of ZP domains in human, mollusk and yeast. This suggests that the overall mechanisms may be similar. In contrast, for sperm egg-binding and entry proteins there is generally no homology between distantly related species. The striking diversity of sperm proteins involved in egg-coat entry reflects the variation of selective pressures and speciation events over a long evolutionary history.

Before the sperm penetrates the egg envelope it must undergo the acrosome reaction (AR), the release of the contents of a vesicle in the sperm head. In marine invertebrates there is variation in the importance of this step in the species-specificity of fertilization. The sea urchins sperm Receptor for Egg Jelly (suREJ1) and egg fucose sulfate polymers interact to induce the AR (Vacquier and Moy 1997). The fucan polymers vary between urchin species and are species-specific inducers of the acrosome reaction (Vilela-Silva et al. 2008). However, their species-specificity and importance as a barrier to hybridization differ across species pairs. Sea stars also have an egg jelly polysaccharide recognition and AR induction system, but it is so far only found to be sub-family specific (Nakachi et al. 2006). A later step, mediated by bindin in sea urchins, may be more important for species-specific fertilization. Receptor suREJ1 has a mammalian homolog, Pkdrej (Hughes et al. 1999), which localizes to the sperm head and is evolving under positive selection in primates (Hamm et al. 2007). However, it appears

to be involved in the timing of the AR in mammals rather than having a direct interaction with the egg (Sutton et al. 2008).

Sperm-egg envelope interactions are best characterized in urchin, abalone, and mouse. The egg envelope is a major hybridization barrier in mammal *in vitro* fertilization; its removal allows greater heterospecific membrane fusion (Yanagimachi 1994). It appears that a complex interaction of the ZP proteins is important for species-specific recognition in mammalian eggs (Yauger et al. 2011). Abalone lysin and VERL was the first pair of interacting fertilization proteins to be identified (Swanson and Vacquier 1997). Lysin creates a hole in the egg vitelline envelope through non-enzymatic interactions with Vitelline Envelope Receptor for Lysin (Lewis et al. 1982). Sperm proteins lysin and sea urchin bindin are both rapidly evolving under positive selection (Lee and Vacquier 1992; Lee et al. 1995; Metz and Palumbi 1996). Egg receptor proteins have more recently been sequenced and found to also be under positive selection. For example, abalone VERL and mammalian ZP2 and ZP3 show positive selection (Swanson et al. 2001; Galindo et al. 2003; Turner and Hoekstra 2006), and urchin EBR1 activity is species-specific (Kamei and Glabe 2003). Other abalone ZP domain proteins of unknown function are also under positive selection (Aagaard et al. 2006). Rapidly evolving residues can often be correlated to ligand-receptor binding domains, implicating selection on the interaction between sperm and egg. Sea urchin bindin has been used extensively to study the evolutionary dynamics of sperm-egg interaction. For example, Palumbi et al. (1999) found that the success of a sperm depends on its and the female's bindin genotype (Palumbi 1999). Levitan and Stapper found that this effect is dependent on population densities (Levitan and Stapper 2010). Lysin and VERL, along with

mammalian sperm sp56 and egg envelope protein ZP3, show evolutionary patterns that indicate coevolution (Clark et al. 2009; Rohlfs et al. 2010).

Sperm-egg membrane fusion mechanisms are less well understood. Known and putative gamete fusagens vary in their sequence conservation. HAP2-GCS1 is a gamete fusagen in *Plasmodium*, *Chlamydomonas* and *Arabidopsis*. It is hypothesized to be the ancestral gene for gamete fusion, due to its presence in many animal genomes, and its expression in the testis of *Hydra*, a basal cnidarian (Wong and Johnson 2010). However, it appears to have been lost in many lineages, suggesting redundant molecular mechanisms for gamete fusion, maybe as a result of positive selection. Positive selection is seen in fusagenic sperm proteins in mammals and marine invertebrates. Sea urchin bindin and abalone sp18 are both acrosome proteins with *in vitro* fusagenic properties, and both evolve rapidly and localize to the acrosome (Swanson and Vacquier 1995a; Ulrich et al. 1998). In mouse, 23 membrane proteins from the sperm surface and acrosomal membrane show evidence of positive selection (Dorus et al. 2010), and in mammals 5 sperm surface ADAM proteins show positive selection in their adhesion domains (Finn and Civetta 2010). Phenotypes in mouse knockout studies frequently contradict *in vitro* results. One current hypothesis is that egg CD9 plays a role in fusion in mice, and Izumo is a candidate interacting protein (for a recent review of mammalian sperm-egg fusion, please see (Rubinstein et al. 2006)). Like other gamete interaction proteins, CD9 has sites under positive selection (Swanson et al. 2003). In *Drosophila* the sperm enters the egg and must go through plasma membrane breakdown (PMBD). Two proteins, Snky and Mfr, are required for efficient PMBD, but their specific functions are still unclear (Wilson et al., 2006; Smith and Wakimoto, 2007). Both appear to have

homologs in distant species, but have not been found in studies of mosquito testis transcriptomes and sperm proteomes (Krzywinska and Krzywinski 2009; Sirot et al. 2011). The prevalence of positive selection on fusagenic proteins may vary based on the fertilization system. If there is selection for a barrier to fertilization and rapid evolution at earlier steps results in successful adaptations, there may be relaxed selection on the later steps.

Another class of proteins exhibiting rapid evolution in internally fertilizing species is seminal fluid proteins, which are ejaculated along with sperm. Ramm et al. (Ramm et al. 2009) found interspecific diversity in seminal fluid protein composition and sequence within muroid rodents. Seminal fluid and sperm protein evolution can also be compared by dissection and partitioning of male reproductive tissues for comparative proteomics and EST sequencing. In mice, seminal vesicle proteins evolve rapidly on average, but proteins from the other male reproductive tract tissues are under evolutionary constraint when compared to the whole genome (Dean et al. 2009). In insects, cross-species studies show that accessory gland proteins have increased evolutionary rates (Wagstaff and Begun 2005; Andres et al. 2006). In field crickets, positive selection is seen in seminal fluid proteins between closely related species (Andres et al. 2006; Marshall et al. 2010), and there appears to be little conservation in overall insect seminal fluid protein composition (Walters and Harrison 2010). Two *Gryllus* species of crickets that have rapidly diverging seminal fluid proteomes, but are otherwise closely related, are an example of how we can use these systems to study the evolution of reproductive isolation in hybrid zones(Andres et al. 2008). The specific functions and interactions of seminal fluid proteins are still largely uncharacterized.

Accessory Gland Proteins in *Drosophila* modulate female post-mating behavior, such as causing an increase in egg laying (Wolfner 2009).

1.3 Causes of rapid evolution

We have long observed the pattern of rapidly evolving gamete interaction proteins in a wide range of taxa (Swanson and Vacquier 2002). Identification of putative sperm-egg interaction proteins under positive selection is becoming easier as new technologies allow faster and cheaper DNA sequencing and protein identification by mass spectrometry. However, testing hypotheses about why these proteins evolve rapidly remains a challenge. What selective pressures are acting in each situation? It is likely that an interplay of forces act in a variable way in each group of species. Here we will discuss the various hypotheses to explain rapid gamete interaction protein evolution, and the data that support or refute each hypothesis.

Some proposed mechanisms for rapid evolution of reproductive proteins deal with the within-species effects such as sexual conflict and sperm competition (Figure 1.2). High population density can drive sexual conflict over the mating rate, which may result in several possible patterns of evolution. Mathematical models demonstrate that runaway coevolution can result in the evolution of reproductive barriers within species (Gavrilets 2000). In one model, females diversify into two distinct groups due to sexual conflict. In response, males may either be stuck in an intermediate fitness state or diversify to match each genotype, resulting in sympatric speciation (Gavrilets and Waxman 2002). Experimental studies of the relationship of sperm densities, the genotype of a protein involved in sperm-egg interaction, and reproductive success are limited to sea urchin bindin. Reproductive success in *Echinometra* sea urchins is dependent on male bindin genotype, and varies according to female bindin genotype (Palumbi 1999). Levitan and

collegues have conducted many studies on the effect of male and female density on reproductive success. In *Strongylocentrotus franciscanus*, there is an optimum range of female reproductive success between sperm limitation and polyspermy, and a relationship between sperm bindin genotype success and density (Levitan and Ferrell 2006). Consistent with predictions of sexual conflict, rare male alleles are more successful in high-density situations. In a high-density living *Strongylocentrotus* species, common bindin alleles were generally more advantageous, but selection maintains some less common, highly successful, variants (Levitan and Stapper 2010). These results would be even more interesting if they were able to also include the EBR1 sequence of the females, and attempt to better understand the evolution of the interaction itself.

We are also beginning to see these patterns in mammals. Sperm competition can affect the development of fertilization barriers between closely related rodents (Martín-Coello et al. 2009). As female proteins that prevent polyspermy are selected for, diversification arises, and a byproduct of that is differentiation between species. Sperm competition may also affect primate seminal fluid genes. Positive selection in a main component of the semen coagulum in primates correlates with higher female promiscuity (Dorus et al. 2004).

Alternatively, the reinforcement hypothesis states that diversifying selection for fertilization proteins to prevent hybridization when spawning or copulation between species overlaps causes rapid evolution (review: (Noor 1999)). This hypothesis can apply to behavioral and ecological factors as well as the proteins involved in direct gamete interaction. The latter is best studied in free-spawning marine organisms. Reinforcement in species such as mussels, abalone, and sea urchin can be tested by

comparing the rates of evolution of a gamete interaction protein between sympatric and allopatric populations. The data produced to date do not consistently support this hypothesis. Single studies have found no evidence for reinforcement in abalone lysin (Clark et al. 2007) or ascidian gamete recognition proteins (Nydam and Harrison 2011). In *Mytilus* species of mussels, Springer & Crespi (Springer and Crespi 2007) found lysin-M7 divergence patterns supporting the reinforcement hypothesis, but in another study Slaughter et al. (Slaughter et al. 2008) tested gamete compatibility and found greater compatibility in sympatry, suggesting reinforcement is not the dominant selective force. Riginos et al. (Riginos et al. 2006) also did not detect reproductive character displacement between sympatric and allopatric species of *Mytilus*. The most data in this field is from studies of sea urchin and its sperm protein bindin. Bindin evolution studies vary in their implications. Two species of South Pacific *Echinometra* have divergence and positive selection in sympatric populations, but some share alleles in allopatric populations, supporting reinforcement (Geyer and Palumbi 2003). However, further studies of *Echinometra* species found bindin positive selection and heterospecific incompatibility in allopatric populations (Metz et al. 1994) (McCartney and Lessios 2004; Geyer and Lessios 2009). These results suggest that some interspecific selective pressures such as sexual conflict or sperm competition may be driving the rapid evolution of bindin. In addition, an understanding of the evolution of the female receptor protein for bindin in each case would also be useful because reinforcement suggests selective pressure should be stronger on the egg, because it is most negatively affected by non-specific fertilization.

1.4 Methods to identify interacting proteins

Sexual conflict and sperm competition drive rapid divergence between sperm-egg interacting proteins. These are potential mechanisms of sympatric speciation (Gavrilets and Waxman 2002). We must identify the proteins responsible for species isolation in order to study past and current speciation events. Identification methods include biochemical, immunochemical, and genetic approaches. Gamete interaction proteins were first identified with biochemical purification, starting with sea urchin bindin (Vacquier and Moy 1977). This is why abalone and sea urchin, in which large amounts of gametes can be easily obtained, were the earliest models of fertilization. Mammalian interactions have been tested with *in vitro* antibody blocking studies, but these may be non-specific in their results. Genetic knockouts in mouse are useful because you can remove a single protein and test the specific result. However, sometimes they give results that conflict with *in vitro* protein function predictions (e.g. (Baba et al. 1994)). One reason could be that the interaction relies on a combination of proteins that remain functional with the loss of one member, or some other form of redundancy. In knockout studies, it is important to also characterize the ability of mutant sperm to compete with wild type. This was key in revealing a phenotype for mutant Pkdrej mice (Sutton et al. 2008). A benefit of using free-spawning invertebrates as model systems is that you can more easily observe fertilization under natural conditions than in those with internal fertilizations such as mammals. Recently, mass spectrometry has allowed the characterization of numerous sperm and seminal fluid proteomes ((Stein et al. 2006) (mouse); (Baker 2007) (human); (Dorus et al. 2006)(fruit fly)). More should be done to describe the proteomes of the egg coat and egg plasma membrane (like (Aagaard et al. 2006; Yamada et al. 2009; Aagaard et al. 2010)).

Detecting putative interacting protein pairs may be possible without first finding clear biochemical evidence. Experimental data have shown that interacting protein pairs often contain signatures of coevolution. Using that signal, Clark et al. (Clark et al. 2009) and Rohlfs et al. (Rohlfs et al. 2010) presented sequence analysis methods to predict protein-protein interaction in reproductive proteins that are physically un-linked. Clark et al. used the known interacting pair lysin and VERL in 8 species of abalone. They showed a correlation of dn/ds values along the branches that were not seen in non-interacting proteins, and determined statistical significance with likelihood ratio tests. Rohlfs et al. detect selection for allele matching using human population genotype data and Composite Linkage Disequilibrium. They found that the human putative interacting proteins egg ZP3 and sperm ZP3R showed allelic associations at more SNPs than those at background genomic levels and random gene pairs. Their result is consistent with the prediction that interacting proteins will fix compensatory mutations as each protein evolves in order to maintain allele matching.

1.5 Conclusions

We have described the current state of knowledge of gamete interaction proteins, their evolution, and methods of analysis and identification. Many more candidate sperm-egg interaction proteins will be identified as DNA sequencing and mass spectrometry become more accessible. Thorough phylogenetic sampling will allow us to assess the diversity and evolution of these candidates. However, it is important to understand their function in gamete interaction with genetic and biochemical studies as well. In addition, we should work to characterize the evolution of more egg proteins, to better understand selection acting on proteins from both sides. Knowing the sequences and binding

domains of interacting protein pairs will be particularly useful for understanding the evolution, divergence, and mechanisms that lead to reproductive isolation.

1.6 Figures

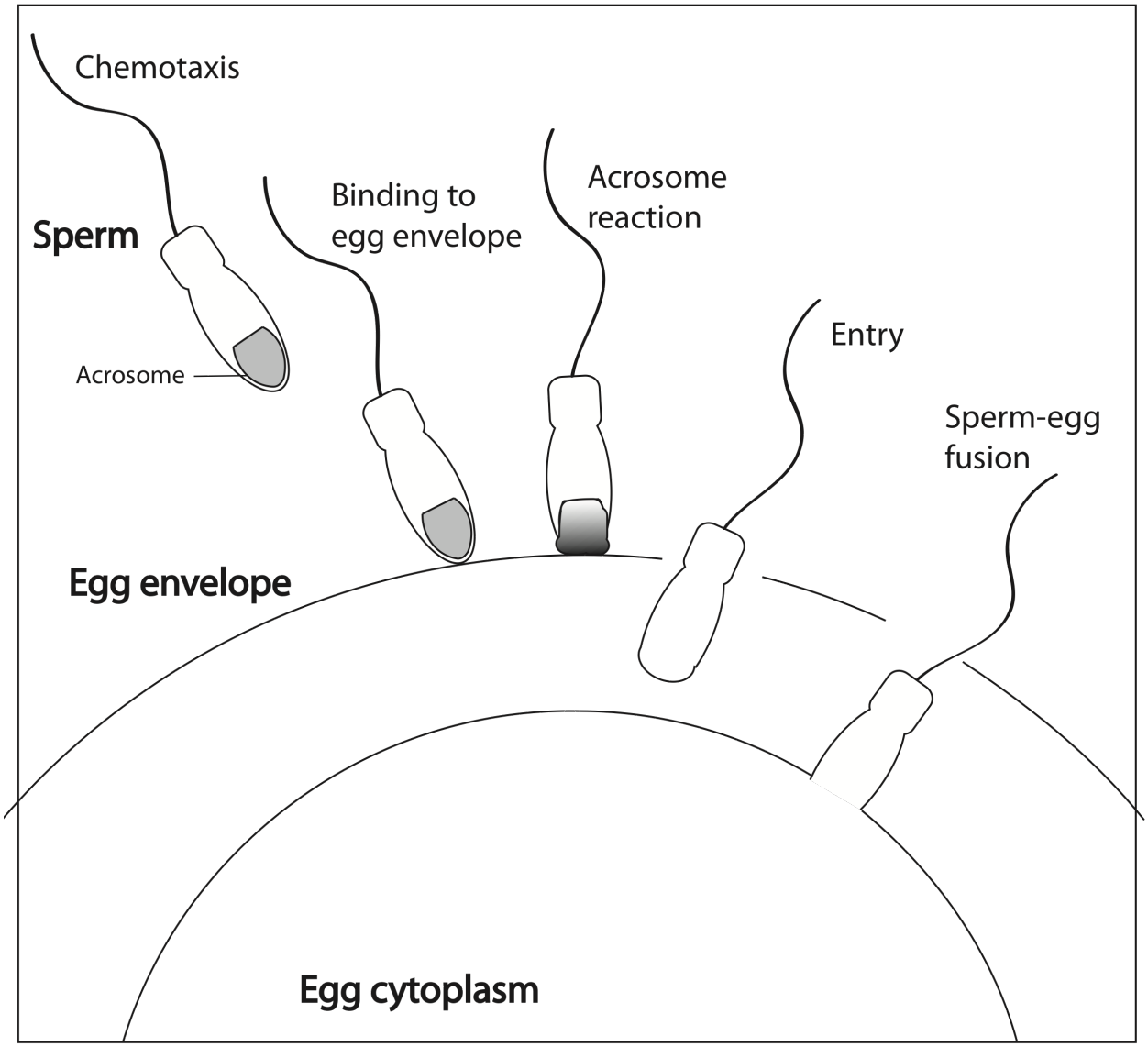


Figure 1.1. A simplified depiction of the basic steps of mammalian and invertebrate fertilization.

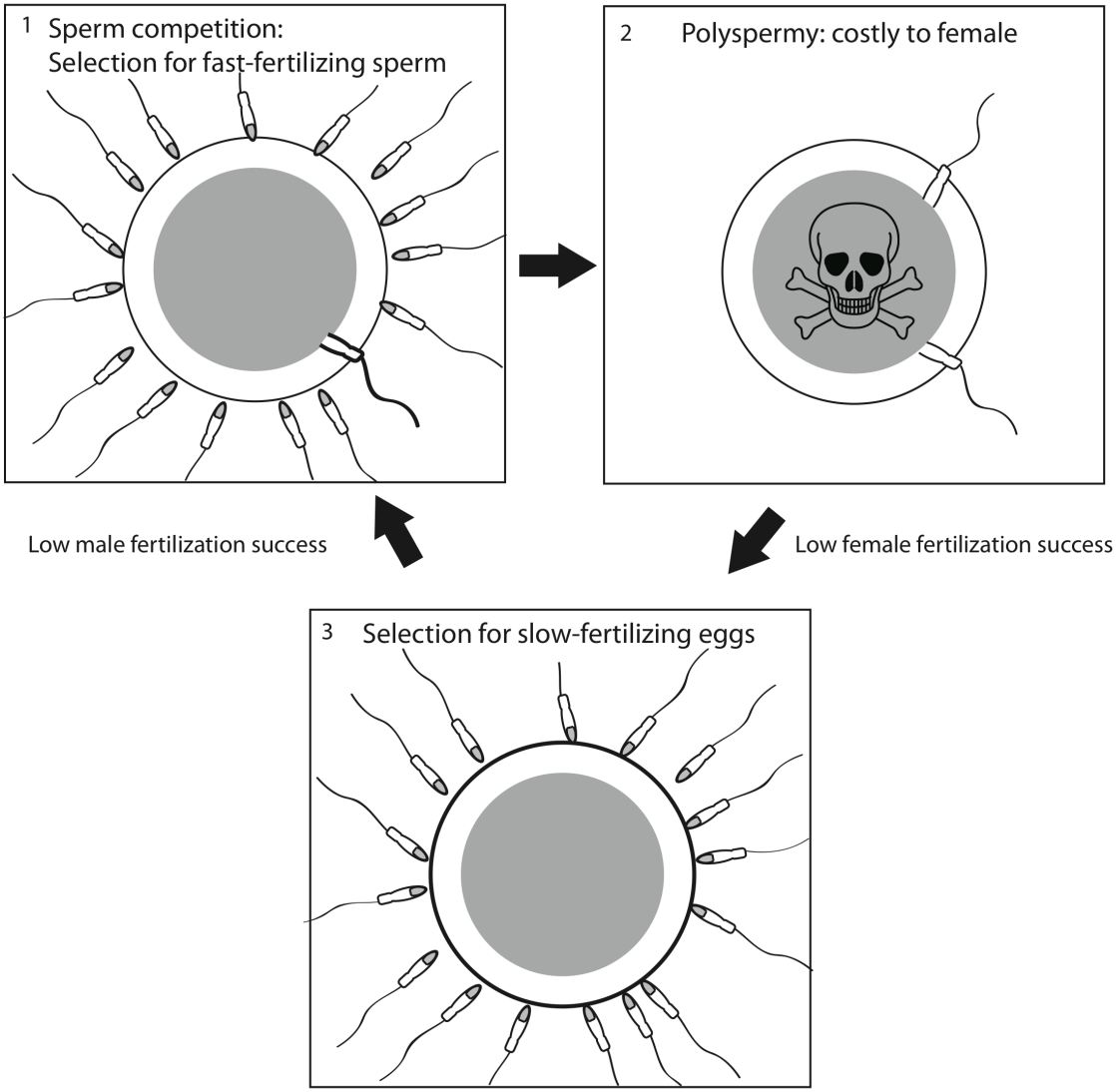


Figure 1.2. A cycle of sexual conflict can drive rapid coevolution between sperm and egg proteins in sperm-dense situations. Sperm competition results in selection for the sperm that can most rapidly fertilize the egg. That may lead to an increase in polyspermy rates, which results in selection for eggs that have stronger blocks to fertilization.

Chapter 2

Mass spectrometry and next-generation sequencing reveal an abundant and rapidly evolving abalone sperm protein

A version of this chapter was previously published as:
Palmer, MR, McDowall, MH, Stewart, L, Ouaddi, A, MacCoss, MJ, and Swanson, WJ. 2013. Molecular Reproduction & Development: Volume 80, Issue 6, pp 460 – 465.

2.1 Introduction

Fertilization is a cascade of events for which surprisingly few molecular interactions are known. A detailed molecular understanding of its mechanisms and evolution begins with the identification of sperm and egg proteins. Abalone, a broadcast spawning marine invertebrate, has a long history as a model system for the biochemical events of fertilization (Swanson and Vacquier 2002), and provides an excellent opportunity for further study. California abalone exhibit species specificity in fertilization as a result of sequence diversity in sperm-egg interacting proteins (Swanson and Vacquier 1998). Abalone sperm have an unusually large acrosome granule (Lewis et al. 1980), which allowed the early purification of the proteins lysin and sp18. Sp18 is implicated in sperm-egg fusion (Swanson and Vacquier 1995b). Lysin dimers create a hole in the egg vitelline envelope, allowing sperm to swim through, by interacting with the Vitelline Envelope Receptor for Lysin (VERL) (Lewis et al. 1982; Kresge et al. 2001). The species-specificity of VE dissolution by lysin is mediated by two highly variable regions in the amino acid sequence (Lyon and Vacquier 1999). The rapid evolution of lysin and VERL may be driven by coevolution between the two proteins (Clark et al. 2009). Lysin in particular has become a textbook example of rapidly

evolving fertilization proteins (Kresge et al. 2001) and was an early focus of study due to its abundance and acrosomal localization. Surprisingly, as demonstrated here, it is not the most abundant acrosomal protein.

The discovery of new abalone sperm-egg interaction proteins is interesting for studying positive selection and important to the elucidation of general mechanisms of fertilization. Most animal egg coats, including abalone's, are made of glycosylated zona pellucida (ZP) domain proteins (Jovine et al. 2005). Thirty Zona Pellucida domain proteins have been identified as components of the abalone egg vitelline envelope in addition to VERL, several of which show evidence of positive selection (Aagaard et al. 2006; Aagaard et al. 2010). Despite rapid evolution in sperm-egg interaction proteins, there is shared ancestry in the structural homology between distant species. For example, abalone VERL repeats are structurally homologous to the ZP-N domain of human egg ZP3 (Swanson et al. 2011). ZP3, like VERL, is under positive selection in mammals (Swanson et al. 2001) and is thought to be involved in sperm-egg coat interactions. Interestingly, the sites subject to positive selection in both molecules fall on the same face in a region implicated in species-specificity, suggesting selection is related to sperm-egg interaction. Lysin and VERL's interactions and evolution are well studied, but there are many other steps in the fertilization cascade that are not characterized at the molecular level. For example, chemoattraction is also species specific between the closely related Red and Green abalone (*Haliotis rufescens and H. fulgens*), but no sperm receptor is known (Riffell et al. 2004).

Broadcast spawning marine invertebrates provide an excellent system to study how selective pressures influence the evolution of reproductive proteins. Sperm-egg

interaction is a major source of diversity in sperm proteins. Aside from abalone, rapid evolution in sperm proteins has been found in sea urchins (*Strongylocentrotus*), top snails (*Tegula*), and ascidian (*Ciona*). In *Strongylocentrotus* species, the sperm protein bindin, which is involved in sperm-egg recognition and binding, is under positive selection (Pujolar and Pogson 2011). In the top snail *Tegula* a protein with no homology to bindin or lysin, TMAP, is the major protein in the acrosome and is under positive selection (Hellberg et al. 2000). In addition, three potential gamete recognition proteins from the sperm of ascidian *Ciona intestinalis* are under positive selection between two closely related lineages (Nydam and Harrison 2011).

We are interested in finding new abalone sperm proteins and candidate sperm-egg interaction proteins. Of particular interest to us are proteins that are abundant, have a transmembrane domain, a signal sequence, or a carbohydrate-interaction domain. Proteins with a transmembrane domain may be involved in the initial egg-recognition event or in membrane fusion, like the gamete fusion protein HAP2-GCS1 (Wong and Johnson 2010). A signal sequence indicates that a protein is secreted, either outside the sperm cell or into the acrosome, and we are interested in identifying acrosome proteins other than lysin and sp18. Egg coats are made up of glycoproteins, so a carbohydrate-interaction domain could be a feature of a protein involved in binding and recognition to the egg coat. In sea urchins, species-specific induction of the acrosome reaction is mediated by sulfated polysaccharides in the egg jelly (Vilela-Silva et al. 2008). Unknown proteins might also be of interest: if they are divergent, they may be evolving under positive selection. In this study we assembled a testis RNA-seq library with no reference sequence, and identified hundreds of sperm proteins with mass spectrometry in the Red

abalone *Haliotis rufescens*. The most interesting result was the discovery of a small, abundant, rapidly evolving pair of homologous proteins. Abalone sperm proteins have been studied for over 20 years, but we have observed these important proteins for the first time by applying new genomic and proteomic technologies to a non-traditional model system.

2.2 Results and Discussion

We identified 975 sperm proteins in the Red abalone, *H. rufescens*, by shotgun mass spectrometry (MS). The most striking result was the discovery of a pair of homologous proteins that appear to be as or more abundant than the major acrosomal proteins lysin and sp18. We can approximate the relative transcript abundance with read coverage of RNA-seq data and protein abundance with the normalized spectral abundance factor (NSAF) (Zybailov et al. 2006) from proteomic MS data. As expected, lysin and sp18 had high coverage in both RNA-seq and MS, but the combined coverage of two novel proteins was higher. One form is approximately 6 kDa, so we call the protein 'sperm protein 6', or sp6. The two forms vary in the length of an aspartic acid stretch along with other differences, so we refer to those as sp6_4D and sp6_8D (4 aspartic acids and 8 aspartic acids, respectively). Both sequences have an N-terminal signal peptide, predicted by SignalP and supported by mass spectra coverage starting at amino acid 20. Mature sp6_8D has a predicted pI of 4.13, MW of 6.86 kDa. Mature sp6_4D has a predicted pI of 3.93, MW of 6.28 kDa. The highly acidic nature of these novel peptides is in stark contrast to the strongly basic nature of lysin and sp18.

We purified sp6_4D and sp6_8D, which had previously not been observed by SDS protein gel electrophoresis. They are very small proteins, so they migrate past the

end of standard gels. In addition they do not stain well with Coomassie, due to their acidic nature (Tal et al. 1985). When we performed SDS-PAGE using 15% acrylamide on sperm or testis proteins and silver stained, a dark band appeared around the predicted sizes of the two sp6 proteins (Figure 2.1A). The first lane shows acrosome-enriched testis proteins, the second acrosome-depleted, and the third ovary proteins. To confirm that the band contains sp6 proteins, we excised the gel bands and performed MS and found that sp6_8D is the dominant protein. We purified sp6 proteins from sperm protein extracts with an anion exchange column for further characterization, and used the product to develop an antibody. Purification and antibody generation were done with a fraction that includes both sp6 proteins, so work with these refers to sp6 in the singular form. We used antibodies to show that sp6 is not present in ovary protein extracts, and is associated with the acrosome of abalone sperm, using Western blots and sperm immunofluorescence (Figure 2.1). The high abundance and acrosomal localization strongly suggests sp6 plays an important role in fertilization.

With PCR using primers from the known *H. rufescens* sequences, we identified *sp6* forms expressed in the testis of four other *Haliotis* species (Figure 2.2). We also looked for *sp6* in two available *Haliotis* whole-body EST databases, using BLAST, with our sequenced *sp6* sequences as queries. We found one homolog in the *H. diversicolor* database (Jiang et al. 2011), but not in the South African abalone *H. midae*. For *H. rufescens*, *H. discus*, and *H. kamtchaskana*, agarose gels of PCR products showed bands at multiple sizes, and, as expected, we sequenced multiple forms from those clones. For *H. cracherodii* and *H. discus* we only amplified one form of *sp6*. The sequences are highly divergent, between forms and between closely related species. BLAST searches

against the NCBI protein database and the gastropod mollusk *L. gigantea* EST database revealed no homologous proteins. Sp6 proteins vary in length, largely determined by a region of variable-length aspartic acid repeats. We don't know the mechanism and consequences of this variation, but it is interesting to note that tandem repeats are present and variable in other rapidly evolving sperm proteins. For example, oyster bindin's F-lectin domain repeats are both polymorphic between individuals and variable within individuals as a result of alternative splicing and recombination (Moy et al. 2008). In addition, indel variation appears to be prevalent in *Drosophila* accessory gland proteins, which are delivered to the female upon mating (Schully and Hellberg 2006).

The high level of sequence diversity made it difficult to align sp6 sequences. High confidence alignments are necessary to analyze substitution rates to detect positive selection, so we created two separate alignments for alignable subsets of sp6_4D and sp6_8D, removing the signal sequence and any gaps, and calculated d_N/d_S with codeml. Both forms of *sp6* show evidence of positive selection among species (Table 2.1). In fact, sp6 appears to evolve more rapidly than either lysin or sp18, suggesting it may play a role in species-specificity.

Sp6 is abundant, evolves rapidly, and may be abalone specific, based on BLAST queries of other species databases. Given its location and rapid evolution, one possibility is that sp6 interacts with lysin and/or sp18. Lysin and sp18 are both positively charged, so sp6 may help them remain soluble while tightly packed in the acrosome. Or it may interact with an egg protein to facilitate species-specific fertilization. Alternatively, it may act as an anti-microbial peptide, protecting the fertilized egg or the developing

embryo from surrounding pathogens. Two anti-bacterial peptides have previously been isolated from whole abalone (Park et al. 2012).

We have also identified many other abalone sperm proteins that may be essential for fertilization and species specificity. We used RNA-seq to generate a transcriptome assembly for *H. rufescens* testis. The de novo assembly yielded 41,300 contigs. We added capillary-sequenced cDNA clones from a testis library to the assembly file for a final contig number of 41,576. We identified 975 total sperm proteins with shotgun mass spectrometry of whole sperm, acrosome enriched fractions, and extracellular digestions. The final assembly and the protein list are available at http://depts.washington.edu/swansonw/Swanson_Lab/Data.html. In a BLAST search to the NCBI non-redundant protein database, 676 of the proteins have a significant match. Our sperm protein numbers are similar to those found in mouse: 858 (Baker et al. 2008), fruit fly: 1108 (Wasbrough et al. 2010), and human: 1056 (Baker 2007), but greater than seen so far in ascidian: 304 (Nakachi et al. 2011). While sp6 is the most striking example of rapid evolution in the newly identified sperm proteins, we also found other proteins that are under positive selection (Table 2.1).

Sp6 was not discovered until we used new genomic and proteomic technologies for protein discovery to learn more about abalone fertilization. The combination of testis transcriptome and sperm proteome gives us an excellent list of candidate fertilization proteins, in addition to sp6. This is particularly useful for abalone, because it contributes more information to an already well-studied system. We will perform more detailed studies of the functions of sp6 and other proteins that may play an important role in abalone fertilization, and combine this information with what we know about the egg coat

proteome (Aagaard et al. 2010). This will advance our understanding of the entire fertilization cascade in a model system, which also provides the opportunity to study how selection can affect evolutionary rates at different steps of fertilization.

2.3 Methods

Strand-specific cDNA was prepared from RNA from the testes of 5 *H. rufescens.* Sequencing was performed on an Illumina Genome Analyzer with 76 base pair paired-end reads. Raw reads are available at the NCBI Sequence Read Archive, accession number SRR770266. The paired reads were filtered with a phred quality cutoff of 5. Assembly was performed with the Trinity package (Grabherr et al. 2011), then contigs were added from Sanger sequencing of a cDNA library for a total of 41,576 predicted transcripts. Read coverage of contigs was estimated with the BWA and Samtools packages.

Spawning was induced in *H. rufescens* males in 0.2μm filtered seawater with the Tris/H_2O_2 protocol (Morse et al. 1977). Intact sperm were treated with trypsin or elastase (Promega) to extract extracellular proteins, or induced to acrosome react by addition of a calcium ionophore (Sigma A23187) to 0.2 mg/ml and incubation on ice until most sperm were acrosome reacted in a subsample viewed with a microscope, or spun down and homogenized intact. All protein preparations were digested with trypsin or elastase. Peptides were eluted over a four-hour gradient through an in-line HPLC column and into an LTQ ion-trap mass spectrometer (Thermo) for tandem mass spectrometry. Mass spectra were searched against a six-frame translation of the transcriptome assembly for protein identification, with a decoy database of reversed sequences to determine the false discovery rate, using Sequest (Eng et al. 1994). Data were visualized and summarized

with the University of Washington's Yeast Resource Center's Mass Spectrometry Data Platform: MSDaPl (Sharma et al. 2012). Relative abundance of proteins in a sample was estimated by the normalized spectral abundance factor (NSAF) (Paoletti et al. 2006).

To determine homology to known proteins, Blast2GO (Götz et al. 2008) was run against the non-human, non-mouse EST database, and the non-redundant protein database in NCBI. The protein and transcript sequences were also searched for transmembrane domains and signal sequences, using HMM (THMM Server v. 2.0, CGS Technical University of Denmark) and SignalP (Petersen et al. 2011). Protein molecular weights and charges were estimated with tools on the ExPASy Bioinformatics Resource Portal. From *H. rufescens* candidate sequences, primer sequences were designed to PCR amplify the transcripts from testis cDNA of other species. d_N/d_S analyses were done with the codeml package in PAML (Yang 1997). These sequences were deposited in GenBank (accession numbers KC752594 – KC752618).

Sp6 was purified with anion-exchange chromatography and confirmed with SDS-PAGE. Polyclonal antibodies were raised against purified sp6 from *H. rufescens* testis in rabbits by R&R Research, LLC (Stanwood, WA), and diluted serum was used for Western blots and sperm immunofluorescence.

2.4 Figures and Tables

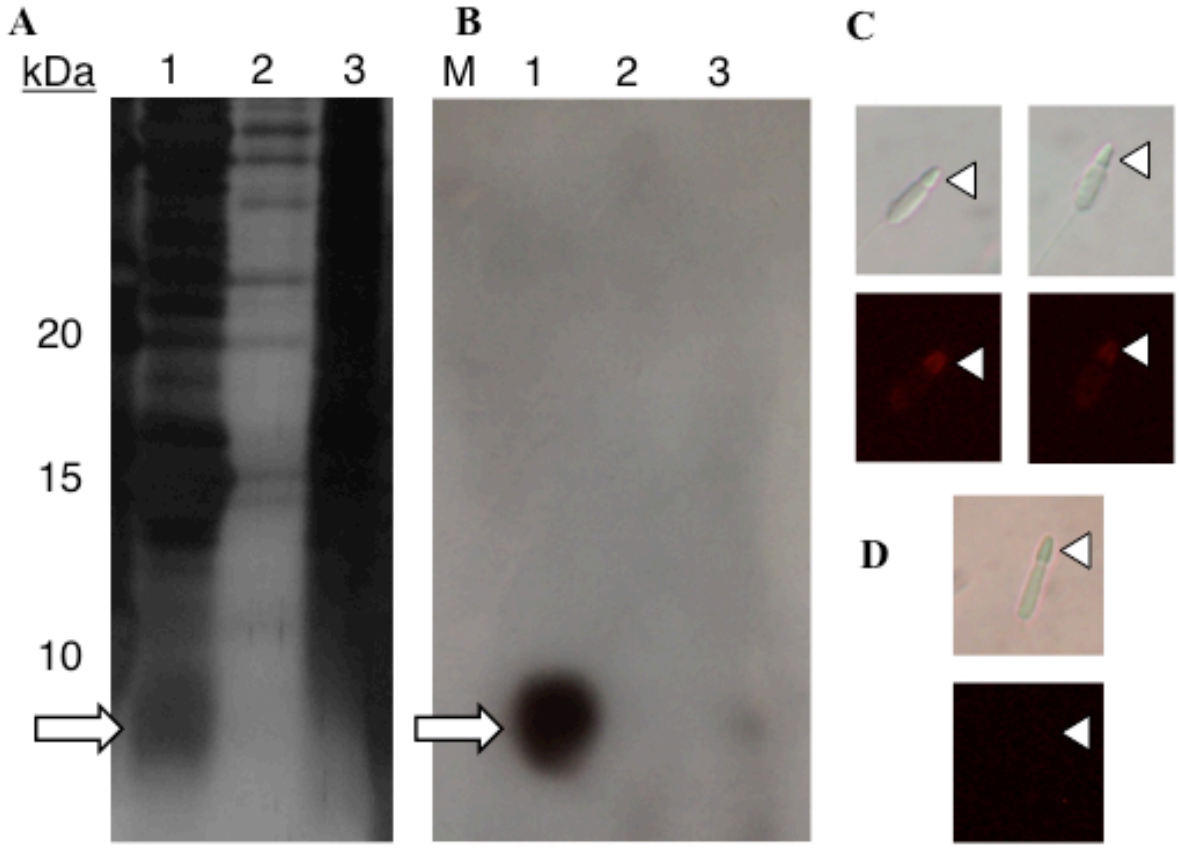


Figure 2.1. Sp6 resides in the acrosome. In A and B, sp6 is indicated with an arrow. In C and D, the acrosome is indicated with a triangle. A)15% SDS-PAGE with sizes marked on far left. Lane 1: acrosome-enriched testis proteins, 2: acrosome-depleted testis proteins, 3: ovary proteins. B) Western blot of the same proteins with sp6 serum. C) Sperm stained with sp6 serum and a rhodamine-conjugated secondary antibody show fluorescence in the acrosome. D) No acrosome staining is seen in sperm incubated with pre-immune serum.

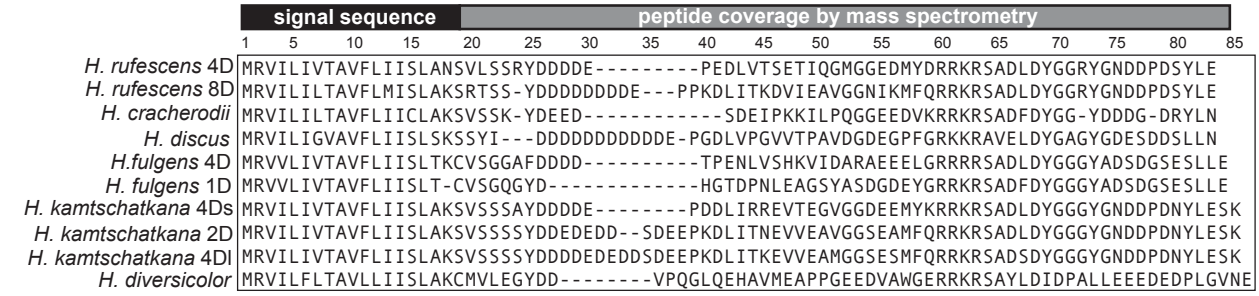


Figure 2.2. Amino acid sequences of sp6 from *Haliotis* species. The gray bar shows sequence that is covered by peptide sequences resulting from shotgun mass spectrometry. The black bar indicates the signal sequence, both as predicted by SignalP and supported by a lack of mass spectra coverage.

Table 2.1. d_N/d_S analysis on abalone sperm proteins, using codeml in PAML. Alignments used for sp6 sequences are available in Appendix A, Figure S2.

Contig Name	Species	d_N/d_S	2ΔlnL M7 v. M8	2ΔlnL pairwise	feature
sp6_8D	*H. rufescens* *H. discus* *H. kamtschatkana 2D* *H. kamtschatkana 4Ds*	5.461	12.335*	sp6_8D	abundant, no blast match
sp6_4D	*H. rufescens* *H. fulgens 4D* *H. kamtschatkana 4Ds*	10.725	17.844*	sp6_4D	abundant, no blast match
comp6_seq1	*H. rufescens* *H. fulgens*	0.417	-	2.187	abundant, no blast match
comp81_seq1	*H. rufescens* *H. fulgens*	0.722		0.518	calcium binding
comp230_seq1	*H. rufescens* *H. discus*	1.172	-	0.779	transmembrane
comp266_seq4	*H. rufescens* *H. discus* *H. walallensis* *H. fulgens*	2.669	0.996		abundant
comp694_seq1	*H. rufescens* *H. discus* *H. fulgens*	3.173	4.99		glycosyltransferase
comp1804_seq3	*H. rufescens* *H. discus* *H. walallensis*	1.640	12.597*		abundant

*: $p < 0.005$, 2 degrees of freedom

Chapter 3

Characterization of a pair of new *H. rufescens* sperm proteins

3.1 Introduction

The molecular details of reproduction have long been of interest from the perspective of both evolutionary and medical research. What proteins interact, how are they influenced by selection, and how can we influence the key proteins for either contraception or to improve fertility? There are now many proteomic profiles of gametes and reproductive tissues, since mass spectrometry has become more accessible and powerful. Lists of proteins are a good starting point, but we must take the next step and figure out the function of newly discovered proteins. This is much more difficult and time consuming, but is important for two reasons. Characterizing the function of an unknown reproductive protein can inform us about the evolutionary strategies for maintaining species specificity in an otherwise conserved process. Also, it can shed light on general biochemical principles of molecular and cellular interactions.

Many new reproductive proteins are being discovered with shotgun proteomics, but few studies have done follow-up functional work. Mammalian reproductive biology is an active field, and many new proteomes have been published, but many processes are still not well understood. One study in humans looked at the proteome of the epididymis, but took it further by raising antibodies to the majority of the proteins and localizing the proteins in epididymal tissue sections and whole sperm (Li et al. 2011). Insects provide the opportunity to study rapid evolution in reproduction, and also to better understand agricultural pests and disease vectors. There are now many published insect sperm and seminal fluid proteomes. Helinski et al. (2012) injected mosquito seminal fluid proteins

(SFPs) into virgin females, and were able to see effects on sexual receptivity (Helinski et al. 2012). Fruit flies provide an excellent opportunity for follow-up work, since genetic tools and RNAi can be used to study the function of newly discovered proteins. *Drosophila* seminal fluid proteins have significant effects on female behavior and physiology, including a reduction in immune defense (Short et al. 2012).

However proteins are identified, one major goal of research in reproductive biology is to find sperm and egg binding partners. Different steps in fertilization show variety in their species specificity and efficiency, and knowing the cognate molecules is key to studying these steps. In mammals, confirming these interactions has been difficult. The most common approach to these problems has been to apply a variety of biochemical techniques to look for binding. The mouse egg coat protein ZP3's proposed sperm binding partner, sp56, was identified by cross-linking and affinity chromatography experiments (Bleil and Wassarman 1990; Li et al. 2010). Later this was supported by co-immunoprecipitation and antibody inhibition in sperm-egg binding assays (Cohen and Wassarman 2001; Helinski et al. 2012). Another study used in vitro expression of target proteins in cell lines to look at binding, and showed that antibodies and other inhibitory molecules could prevent the cell adhesion of cells expressing sperm ADAM and egg integrin proteins from mice (Tomczuk et al. 2003; Short et al. 2012).

Marine invertebrates are another source of model systems for understanding sperm-egg interaction. The egg receptor for sea urchin sperm bindin was identified with radiolabeling and binding experiments (Glabe and Vacquier 1978). The marine mollusk abalone has long been a model for the discovery and characterization of sperm and egg proteins. Abundant gametes can be collected from these broadcast-spawning organisms.

Lysin was the first sperm protein characterized from abalone. It dissolves the egg envelope (Lewis et al. 1982). The receptor for lysin was discovered by affinity purification from a mixture of vitelline envelope proteins, and shown to bind lysin with fluorescence polarization (Swanson and Vacquier 1997). More recently, Aagaard et al. (2010) used the abalone vitelline envelope proteome and affinity columns to determine what VE proteins bind to sp8 and lysin (Aagaard et al. 2010).

Using mass spectrometry to perform shotgun proteomics of abalone sperm, we have discovered a pair of new abalone sperm proteins, sp6_4D and sp6_8D, that are small, abundant, localized to the acrosome, and evolving rapidly between species (Palmer et al. 2013). Based on their location and rapid evolution, we hypothesize that they are involved in sperm egg interaction and species specificity. They may interact with lysin and/or sp18 in the acrosome. I used affinity columns to look for sperm and egg binding partners to the sp6 proteins. I also set up a bacterial expression system for the two separate sp6 proteins, as well as a control protein with similar size and charge. I used high-resolution mass spectrometry and Native PAGE to take a closer look at the sp6 proteins, and look for evidence of modifications. The data from these experiments and the expression system provide an excellent foundation for further studies of the sp6 proteins.

3.2 Results and Discussion

Characterization of sp6 with mass spectrometry

Mass spectrometry of protein mixtures and purified proteins can provide evidence for modifications, show what the mature protein is, and give us an idea of the relative abundance of proteins in a sample. We observed sp6 by mass spectrometry in two ways, with three types of samples: high resolution mass spectrometry using an undigested

sample, purified from testis using anion-exchange, and shotgun mass spectrometry on acrosome-enriched and extracellular-enriched protein samples, both prepared from spawned sperm.

When sp6 was purified by anion exchange, SDS-PAGE gels showed a smaller band than expected by the prediction of the mature protein and by the size of the band on a Western blot of testis proteins (Palmer et al. 2013) (Figure 3.1). Mass spectrometry on undigested anion exchange purified sp6 proteins yielded mostly spectra representing peptides from two fragments of the predicted sp6 protein. Peptide coverage was also fragmented in sp6_8D for all samples, and sp6_4D for the acrosome enriched sample (Figure 3.2). The theoretical pI and molecular weights of these fragments were: sp6_4D: 3.39 and 3183.24 Daltons, sp6_8D: 3.86 and 4058.31 Daltons, and the C-terminal fragment with the same sequence between sp6_4D and sp6_8D: 3.84 and 1557.55 Daltons. In the acrosome and extracellular enrichments, there were more spectra covering the sp6_8D peptides. The purified sp6 samples, however, had more spectra covering the sp6_4D peptides. In both sets there were significantly fewer spectra for the C-terminal peptide (Table 3.1).

There were a few peptides in the anion-exchanged sp6 samples that were not derived from sp6, but they were only represented by one spectra each. Most of them matched to predicted protein sequences that were fragmented or unlikely to be real open reading frames. Two matched to sequences with significant blast hits: a transposase and a Rho GTPase-activating protein.

Alternative isoforms of sp6

Sp6_4D and sp6_8D were the most abundant forms of sp6 in mass spectrometry and cDNA sequencing. However, we saw peptides representing a variant form of sp6_8D in acrosome and extracellular protein enriched samples. They were identical to sp6_8D except for the length of the aspartic acid (D) stretch. There were peptides that span aspartic acid repeat lengths of 7, 9, 10, 11, and 13. Some were represented by more spectra than others (Table 3.2). In some cases, spectra matching the same peptide have different charges, which is also true of sp6_4D and sp6_8D peptides. The cDNA sequences of these proteins are also present in the testis transcriptome. We PCR-amplified and cloned sp6 from cDNA, using various primers, but none of the sequenced clones represented any of these forms.

Expression of sp6 variants with and without V5 tags

Bacterial expression of target proteins allows the synthesis of large quantities of proteins, as well as the potential to create recombinant proteins. We made expression constructs for sp6_4D and sp6_8D, as well as a scrambled form of sp6_4D for use in negative controls, each with two versions: unlabeled and tagged with an N-terminal V5 epitope (referred to as sp6_4Dx, sp6_8Dx, sp6_Scrx, or sp6_4Dx_V5, etc.). For each construct, a small band could be purified from soluble cytoplasmic proteins by anion exchange, with the same method as purification from testis proteins (Figure 3.1). For V5 tagged proteins, we also used a V5 purification kit. V5 purification yielded several size products (Figure 3.3). Only the sp6_8Dx_V5 product reacted with the sp6 serum, and it was around the predicted size of the target protein (8.4 kDa) (Figure 3.3B). A Western with a V5 antibody showed products from all three constructs around 3.5 kDa, which could correspond to the same product seen with anion exchange of expressed or testis

derived anion-exchange purified sp6 (Figure 3.3C). In addition, a larger than expected product appeared, between 10 and 15 kDa, for all three constructs, but had the strongest signal in sp6_4Dx_V5 and sp6_Scrx_V5.

We performed mass spectrometry with the purified product from anion exchange for each construct, labeled or and unlabeled. We saw peptides for only three of the six constructs: sp6_4Dx, sp6_8Dx_V5, and sp6_Scrx_V5. None of them covered the whole protein, and only the sp6_8Dx_V5 peptides covered a fragment of the V5 tag (Figure 3.4). This was likely due to low protein concentrations in the samples.

Anion-exchanged sp6 from testis and expressed proteins often appeared as one blurred or two adjacent bands on a protein gel, around 3.5 kDa. This could be a result of fragmentation of the protein, either by some natural instability, or an enzyme targeting a specific cleavage site. The 'RRKR' residues look like a furin cleavage site, which is also used to make the mature protein in other reproductive proteins, including a rapidly evolving sperm protein from teguline gastropods (Hellberg et al. 2000). However, while the scrambled version of sp6_4D also showed a small product after anion-exchange, it does not have this cleavage site.

The small, anion-exchanged products from testis or expressed sp6 have never shown a signal on a Western blot with sp6 serum. This is surprising because the serum was made by injection of anion exchanged purified sp6 from testis and sperm. The small products did appear in the V5 Western blot, suggesting that they represent the N-terminal end of the protein. This is consistent with the size of the fragment with consistent peptide coverage in mass spectrometry. It is not yet clear which portion of the sequence the serum is reactive to, but this result suggests that it is not the N-terminal fragment. It could

be that an intact protein is required for reactivity, or that the C-terminal end is the key portion.

Non-denaturing gels

If a protein is present in a complex or an aggregate, its progress on a non-denaturing polyacrylamide gel will be retarded. Testis proteins on a non-denaturing polyacrylamide gel with Western blot showed sp6 as a small band, close to the dye front (Figure 3.5). This suggests that the sp6 proteins are not bound to any other proteins in the sperm. However, lysin and sp18 do not migrate on a non-denaturing gel, due to their positive charges. Therefore, it is possible that some sp6 could be bound to them and not run on the gel. Soluble cytoplasmic proteins from the expression constructs for sp6_4Dx and sp6_8Dx show a much bigger product on a non-denaturing gel. This could be due to non-specific antibody hybridization, sp6 being stuck in a complex or aggregate, or the failure of protein synthesis to stop at the end of the sp6 construct in the bacteria.

Affinity columns

To look for proteins that interact with the sp6 proteins, we used affinity chromatography. We linked sp6 proteins to Affi-Gel, and looked at what proteins bound to the column from ovary, zona pellucida, and testis extracts. For two experiments, we used sp6 purified from testis as the ligand bound to the Affi-Gel. We used two control columns: one with no protein bound to the Affi-Gel, and another with a synthesized peptide: a scrambled version of the sp6_4D sequence. SDS-PAGE on eluates from testis protein extracts showed no proteins and no differences between test and control columns.

For ovary and VE extracts, we performed mass spectrometry on eluates containing proteins as indicated by BCA assay and SDS-PAGE. Each sample contained

peptides from around 6,000 predicted proteins, using a 6-frame translated database from *H. rufescens* ovary cDNA sequencing (Aagaard et al. 2010). Only about 50% of protein IDs overlapped, even between similar samples. A total of 14,811 predicted proteins were identified in 9 control and sp6 samples. For detailed analysis, we looked at only the 50 most abundant proteins from each sample, by normalized spectral abundance (NSAF) (Zybailov et al. 2006). From that set, 102 proteins were unique to one or more of the sp6 columns. Four of those proteins were in more than one sp6 column. Of those four, only one had a significant blast match: Calsyntenin-1, a transmembrane calcium binding protein. Another two proteins had single peptides that were covered with many spectra. One was represented by 177 spectra, another by 169 (Table 1 in Appendix B).

We repeated affinity column experiments with expressed proteins as ligands: sp6_4Dx, sp6_8Dx, and controls sp6_scrx and buffer alone. SDS-PAGE showed no clear differences were seen between control and sp6 testis protein columns. Columns incubated with VE proteins showed more protein in the eluates, but no clear differences from controls.

3.3 Future Directions

This work has provided a basis for future functional studies of the sp6 proteins. Many challenges and questions remain. Methods for the purification and stabilization of intact sp6 could still be improved. Sp6 proteins degrade rapidly in samples, and the anion-exchange product is particularly sensitive to degradation, even in the presence of protease inhibitors. According to mass spectrometry, peptides from both sp6 forms are present in those purified samples. While expressing the separate forms is a viable alternative, it would be useful to be able to purify them separately from testis proteins. In

addition, we would like to see if we can separately purify the C-terminal peptide, in order to test its function and binding independent of the N-terminal sequence.

It appears from shotgun proteomics that sp6_8D is more abundant than sp6_4D, but the opposite was true in purified samples. It is not clear whether these differences are biologically significant, or due to variation in how they are detected by mass spectrometry. In addition, we would like to understand the functional significance of the aspartic acid repeat variants of sp6_8D. It would also be useful to know which portion of the protein the sp6 serum reacts to. The expressed proteins could be used to make antisera reactive to specific forms and parts of the proteins.

It would also be interesting to know if the sp6 proteins have any structure. Sp6 is a small protein, so purified sp6 proteins from expression constructs could be used in nuclear magnetic resonance (NMR) to look for any signs of structure.

Despite initial efforts, we still don't know what proteins interact with sp6, or what sp6's function may be in fertilization. For protein interactions, the V5 tagged sp6_8Dx (Figure 3.3) will be useful for co-immunoprecipitation experiments, and continued work should be able to generate an sp6_4Dx version as well. In addition, affinity columns could be made using hypothesized binding partners as the ligand, to see if purified or expressed sp6 will bind in the opposite direction to the affinity columns already done.

Fertilization assays are challenging, but could provide important insights into sp6 function. With fresh sperm and eggs, we could add sp6 antibody to see if it will block fertilization. Also, we could inject sp6 forms into eggs and observe the outcome. If furin cleavage is an important part of the function of sp6, addition of a furin inhibitor to sperm and eggs may change its function and have some effect on fertilization.

Knowing the role of the sp6 proteins in sperm function could contribute to the study of the mechanisms of species-specificity in abalone fertilization. Lysin and sp18 are already known to be under strong Darwinian positive selection. Sp6 can also be used to study evolution within populations and between species, to better understand the process of speciation and the prevention of hybridization in broadcast spawning organisms.

3.4 Methods

Purification of sp6 from testis

Testis tissue from farm-raised red abalone was dissected into 50mM ammonium bicarbonate, 10mM NaCl (or Tris and NaCl), pH 7-8, with or without 0.1% Triton X-100 (about 10ml for 3 animals). Tissue was then homogenized with at least 10 strokes in a glass tissue grinder. Non-detergent preps were also sonicated with a probe set at power level 3, 30 seconds on, 30 seconds off, 5 times. Samples were then centrifuged at 4,000 rpm, at 4ºC, for 10 minutes. The supernatant was moved to a new tube, then centrifuged again at 10,000 rpm for 1 hour. The supernatant of that was used for anion-exchange to purify sp6 proteins.

For anion-exchange, we prepared DEAE cellulose, stored in 1M NaCl solution. 5 ml cellulose was packed into a chromatography column, rinsed with 25ml DI water, 10ml 10x buffer (dissection buffer), and 35ml 1x dissection buffer. Next, protein extract was passed over the column, then the column was washed with 50ml dissection buffer. Proteins were eluted with a step gradient of increasing salt from 0.1M to 1M: 6-8 fractions of 4ml each. Results were assessed with SDS-PAGE, comparing original proteins, column flow-through, wash, and eluates.

Mass spectrometry of un-digested, anion-exchanged sp6

An LTQ-FT Ultra mass spectrometer (Thermo) was used to analyze samples of anion-exchange purified, undigested testis sp6, hoping to find modifications, more instances of the sp6_8+D forms, and any cleavage products. The same was done with anion-exchange purified bacterially expressed sp6 proteins, both V5 tagged and untagged.

Expression of sp6 proteins

Expression constructs were designed with or without N-terminal V5 tags, for the predicted mature sequences of sp6_4D, sp6_8D, and a scrambled version of 4D (sp6_Scrx). They were ordered from IDT in pIDTSMART plasmids containing Kanamycin resistance. Plasmids were transformed into TOP10 cells for storage and propagation. Plasmids isolated from TOP10 cultures were extracted with a Qiagen miniprep kit, digested with NcoI and BamHI, and the cut insert was gel purified from 1% agarose. A pET11d vector (Novagen) was digested with the same enzymes, and purified with a Qiagen PCR purification kit. pET11d and the inserts were combined in a ligation reaction with T4 DNA Ligase (NEB). The ligation products were transformed into NovaBlue cells (Novagen), and selected with Ampicillin. Colonies and ligation reactions were tested with PCR, using combinations of gene-specific and pET11d-specific primers to confirm plasmid identity. Plasmids were then grown up and extracted for sequencing. Colonies with the correct plasmid sequence were stored as glycerol stocks, and plasmids were extracted for transformation into the expression cells, Rosetta-gami 2(DE3) (EMD Millipore). Transformants were selected for with chloramphenicol and ampicillin plates. Colonies were tested with PCR again, sequenced again, and grown in LB with 1% glucose. Clones with the correct sequence were stored as glycerol stocks. For expression,

glycerol stocks were picked into 3ml cultures (50 μg/ml carbenicillin and 34 μg/ml chloramphenicol), grown overnight, used to inoculate 100ml cultures. Those cultures were grown for 2-4 hours, until OD_{600} was close to 0.5, then were split. 75ml of the culture was induced with 100μm IPTG, the rest was left untreated, and they were both grown for 4 or more hours before proteins were extracted from cells. Expressed sp6 proteins appeared to be abundant in the soluble cytoplasmic fraction, and were purified with anion-exchange exactly as from testis proteins.

Affinity columns

For affinity columns, purified sp6_4Dx, sp6_8Dx, sp6_Scrx, or purified sp6 from testis were separately bound to 2ml Affi-Gel® 15 Gel (Bio-Rad). Affi-Gel was washed with water, incubated with sp6 protein or control for 4 hours or overnight, blocked for 1 hour with 0.1M ethanolamine pH 8, and washed with ammonium bicarbonate buffer. Ligand-bound gels were incubated with freshly extracted red abalone testis, ovary, or vitelline envelope proteins and 0.1% Triton X-100 overnight. All incubations were performed at 4°C. Gels were washed, and proteins eluted with 1M glycine pH 3.2 and then 1M NaCl. Eluates were collected and used for SDS-PAGE. 15% acrylamide gels were used for testis proteins, and 2.5-15% acrylamide gradient gels were used for VE and ovary proteins. Eluates from testis sp6 columns, incubated with ovary or VE, were prepared for mass spectrometry by trypsin digestion, as described in Chapter 2.

3.5 Figures

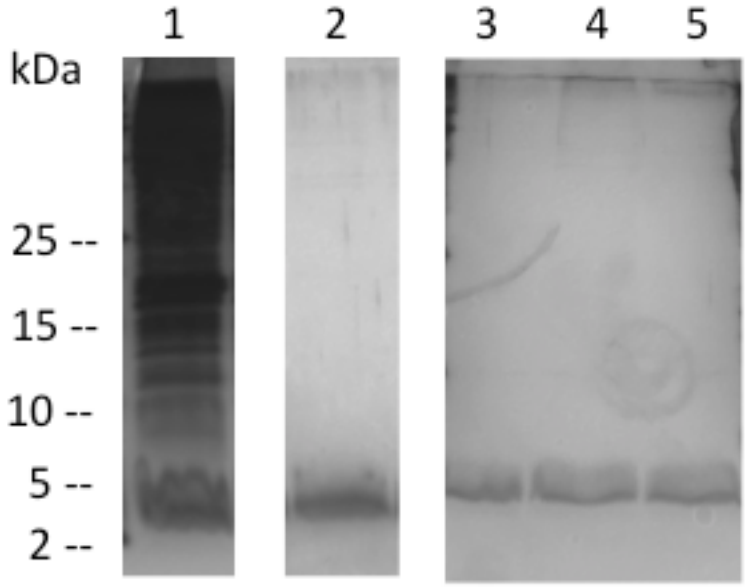


Figure 3.1. 15% SDS-PAGE of 1. Testis proteins, 2. Anion-exchange purified sp6 from testis proteins, and expressed sp6 proteins 3. sp6_4Dx, 4. sp6_8Dx, 5. sp6_Scrx

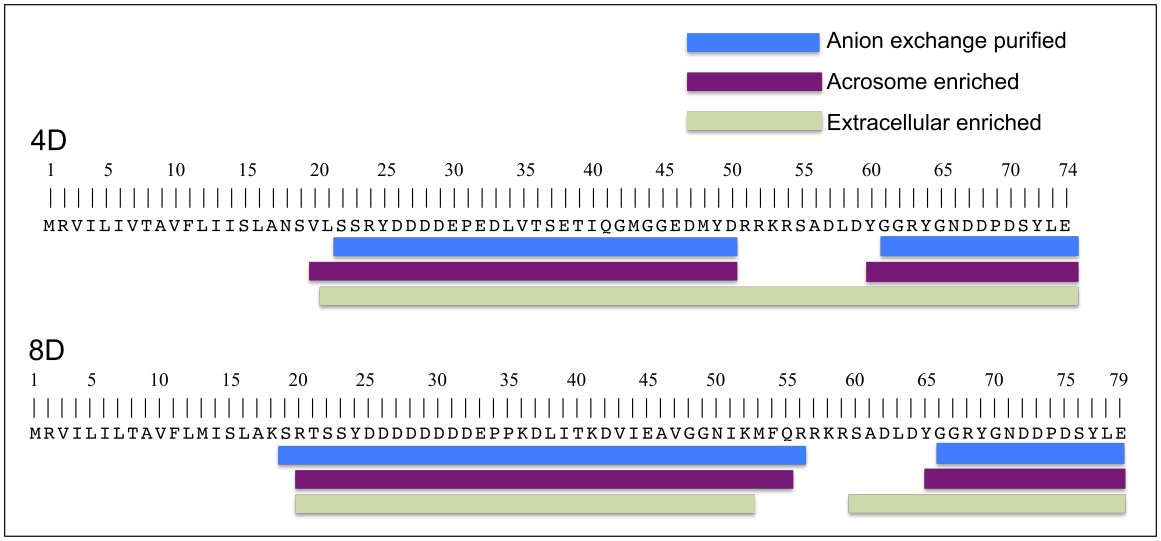


Figure 3.2. Peptide coverage of sp6 proteins sp6_4D and sp6_8D by mass spectrometry of anion exchange purified sp6, acrosome protein enrichment, and extracellular protein enrichment.

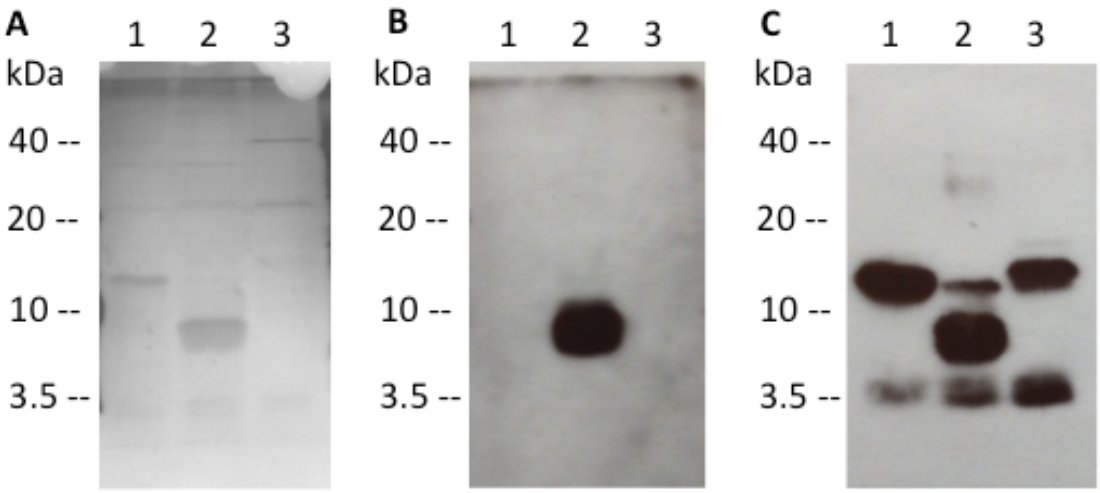


Figure 3.3. V5 purification of V5-tagged sp6 constructs, lane 1: sp6_4Dx_V5, lane 2: sp6_8Dx_V5, lane 3: sp6_scrx_V5. A) 15% SDS-PAGE, B) Western blot with sp6 serum diluted 1:10,000, C) Western blot with Anti-V5-HRP antibody, 1:10,000.

sp6_4Dx:
ARAMGVLSSRYDDDDEPEDLVTSETIQGMGGEDMYDRRKRSADLDYGGRYGNDDPDSYLE
Total spectra: 4
sp6_8Dx_V5
ARAMGKPIPNPLLGLDSTGRTSSYDDDDDDDDDEPPKDLITKDVIEAVGGNIKMFQRRKRSADLDY
GGRYGNDDPDSYLE
Total spectra: 18
sp6_scrx_V5
ARAMGKPIPNPLLGLDSTGSAPLLSDDEVDRGRLSVTSGGISDEPQYDTDDRERNDGDMGEDGKY
SYELYRMDYD
Total spectra: 9

Figure 3.4. Peptide coverage of expressed proteins, highlighted in yellow and underlined.

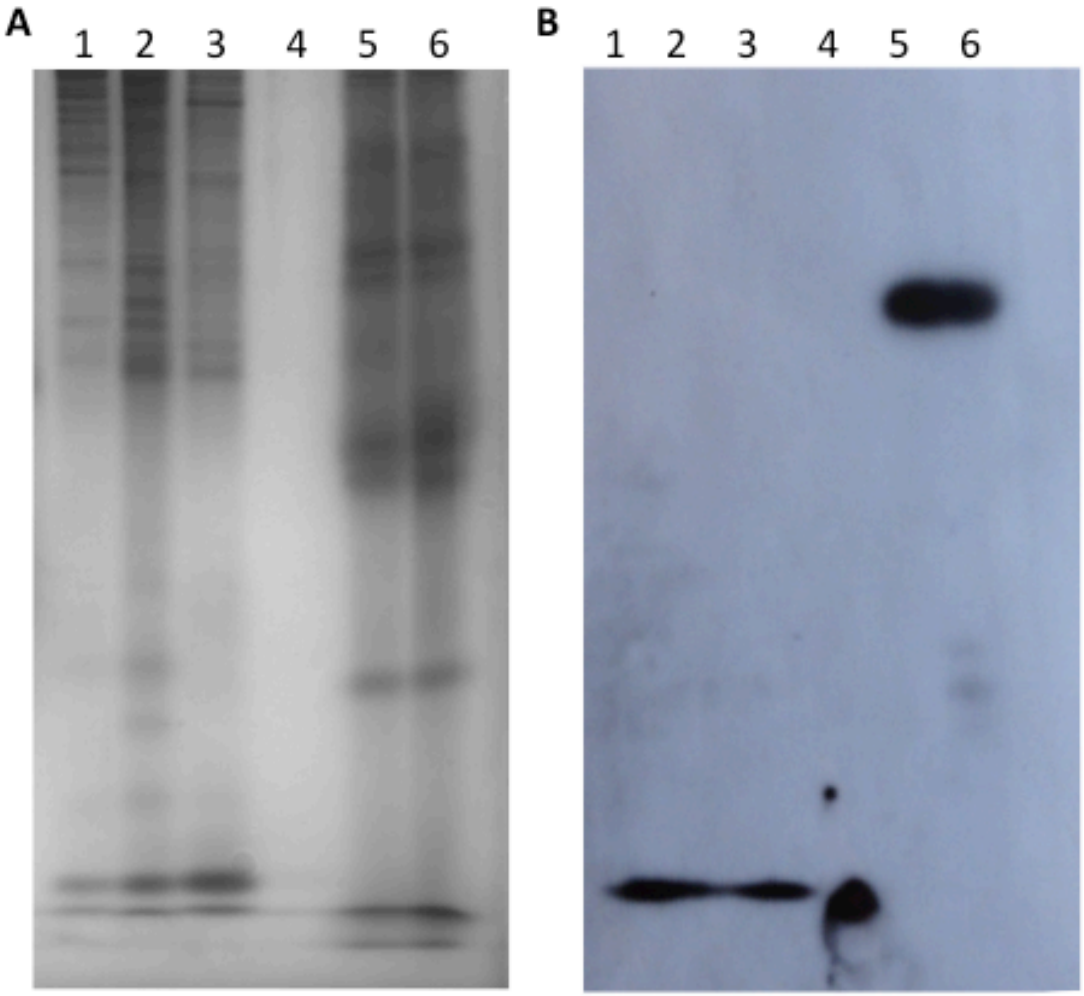


Figure 3.5. A) Native PAGE and B) Western blot with sp6 serum, 1:10,000. Lanes: 1: acrosome enriched testis proteins, 2: testis proteins, 3: Anion exchanged testis proteins, 0.4M NaCl eluate, 4: Anion exchanged testis proteins, 0.6M NaCl eluate, 5: Soluble cytoplasmic proteins from bacteria with sp6_4Dx construct, 6: Soluble cytoplasmic proteins from bacteria with sp6_8Dx construct.

3.6 Tables

Table 3.1. Average number of spectra at each amino acid representing protein-specific peptides.

Protein	Anion-exchanged sp6	Acrosome	Extracellular
4D	61	42	1066
8D	22	559	1789
C-terminal peptide	4	12	154

Table 3.2. Out of 6 biological samples, the number of spectra representing peptides that covered the aspartic acid repeat and at least one flanking amino acid.

D-count	Spectral Count
7	3
9	54
10	1
11	64
13	3

Chapter 4

Summary, Conclusions, & Future directions

This body of work has provided many new directions to study the evolution and function of abalone fertilization proteins. The Red abalone testis transcriptome and sperm proteome revealed many targets for future study. In this portion of the work, we showed the feasibility of proteomic discovery in a species with no genomic sequence resources, even lacking a closely related genome. As computational methods for de novo assembly improve, and as next generation sequencing becomes more affordable and yields longer reads, more model systems that have previously not been amenable to genome sequencing will be able to enter the genomic era.

The most striking discovery from the sperm proteome was the small, abundant protein pair 'sp6.' Many features of the sp6 proteins inform hypotheses about their function. They are small and very negatively charged, as well as being highly abundant in the acrosome. Their charge could be important in balancing out the positive charge of lysin and sp18 in the acrosome, to maintain their solubility. This would also be consistent with their abundance, since lysin & sp18 are the other two major acrosomal proteins. If this were true, we would expect changes in the number of aspartic acid repeats in sp6 to correlate with variation in the charges of lysin or sp18.

So far it is unclear whether sp6 proteins interact with sperm proteins, egg proteins, or both. We should continue to work to find egg interacting proteins for sp6. Sequencing the genes for these proteins in multiple species could reveal variation in copy number and charge that corresponds to that of sp6. Once we can test the function and

interactions of these proteins, we can express recombinant forms of sp6 to test hypotheses about how variations in charge, aspartic acid repeat number, and sequence affect function. One possibility is that the egg receptor has allelic variation, and having multiple forms allows the sperm to interact with more types of eggs. Egg protein evolution could be driving the diversification that we see in sp6.

The sp6 proteins have a very interesting evolutionary history. Many species have at least two copies or alleles, while a few others appear to only have one. Mass spectrometry shows that sp6_8D has five variants with only a change in the number of aspartic acid repeats. Southern blots and sequencing of the genomic region of sp6 should be done to determine the actual gene copy number. Are these tandem repeats caused by unequal crossing over? Are some just allelic variants? Other small proteins like antimicrobial peptides often exist in multiple copies near each other in the genome, and here we could be seeing a similar phenomenon. It appears that they are not highly expressed, so perhaps they are undergoing duplication but not subject to selection. These duplicates could later become advantageous and be under positive selection. We should work to understand the genomic architecture of the sp6 genes within multiple *Haliotis* species, as well as in other genera if they are more widespread.

BLAST searches of the NCBI protein and DNA database, as well as the *Lottia gigantea* genome, show no matches for sp6 in other species. However, representation of closely related species in the database is poor. In addition, the length and diversity of sp6 sequence would make it difficult to find by sequence similarity. We should take a more direct approach to figure out the phylogenetic representation of sp6 proteins. There are several methods that could be used. PCR using primers from the non-coding region of

sp6 was able to amplify sp6 in all *Haliotis* species, and could work in closely related species. It's possible that sp6 and similar small sperm proteins have been missed in many species due to their size and lack of similarity to known proteins. A comparative study of acrosome proteins from broadcast spawning marine invertebrates could look for small, negatively charged proteins.

An alternative hypothesis for sp6 function is that it is an antimicrobial peptide. Antimicrobial peptides are often small and negatively charged. They are subject to proteolytic cleavage, derived from a larger protein, contain a signal sequence, and have a high level of sequence diversity between species, all features of sp6 (Zasloff 2002). Benkendorff et al. (2001) found that antimicrobial peptides are often found surrounding invertebrate embryos in early development (Benkendorff et al. 2001). *Haliotis discus discus* contains two known proteins with antibacterial properties, Nod1 and Nod2, isolated from whole body protein extracts (Park et al. 2012). They are both just over 6 kDa, but only 35 N-terminal amino acids were sequenced. The partial sequence indicates that they are closely related to one another, but have no significant BLAST matches. Our one piece of evidence that sp6 proteins are not antimicrobial is that we are able to express them in *E. coli*, for which the Nod proteins have some inhibitory activity. A more careful experimental approach should be taken to rule out or support this hypothesis.

The work reported here shows that important things can be revealed when we apply new technologies to well-studied biological systems. However, the work to follow up on those new findings can be slow and challenging. Our discovery of the sp6 proteins has provided an opportunity to learn more about abalone fertilization, from both a functional and an evolutionary perspective. We have created a solid foundation to start

this work, and expect that future work will yield very interesting results.

Bibliography

Aagaard JE, Vacquier VD, MacCoss MJ, Swanson WJ. 2010. ZP domain proteins in the abalone egg coat include a paralog of VERL under positive selection that binds lysin and 18-kDa sperm proteins. Molecular Biology and Evolution 27:193–203.

Aagaard JE, Yi X, MacCoss MJ, Swanson WJ. 2006. Rapidly evolving zona pellucida domain proteins are a major component of the vitelline envelope of abalone eggs. P Natl Acad Sci Usa 103:17302–17307.

Andres JA, Maroja LS, Bogdanowicz SM, Swanson WJ, Harrison RG. 2006. Molecular evolution of seminal proteins in field crickets. Molecular Biology and Evolution 23:1574–1584.

Andres JA, Maroja LS, Harrison RG. 2008. Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. P R Soc B 275:1975–1983.

Baba T, Azuma S, Kashiwabara S. 1994. Sperm from mice carrying a targeted mutation of the acrosin gene can penetrate the oocyte zona pellucida and effect fertilization. Journal of Biological ….

Baker MA, Hetherington L, Reeves GM, Aitken RJ. 2008. The mouse sperm proteome characterized via IPG strip prefractionation and LC-MS/MS identification. Proteomics 8:1720–1730.

Baker. 2007. Identification of gene products present in Triton X-100 soluble and insoluble fractions of human spermatozoa lysates using LC-MS/MS analysis. Proteomics Clinical Applications:1–9.

Benkendorff K, Davis A, Bremner J. 2001. Chemical defense in the egg masses of benthic invertebrates: An assessment of antibacterial activity in 39 mollusks and 4 polychaetes. J Invertebr Pathol 78:109–118.

Bleil JD, Wassarman PM. 1990. Identification of a ZP3-binding protein on acrosome-intact mouse sperm by photoaffinity crosslinking. P Natl Acad Sci Usa 87:5563–5567.

Branscomb A, Seger J, White RL. 2000. Evolution of odorant receptors expressed in mammalian testes. Genetics 156:785–797.

Burnett LA, Xiang X, Bieber AL, Chandler DE. 2008. Crisp proteins and sperm chemotaxis: discovery in amphibians and explorations in mammals. Int J Dev Biol 52:489–501.

Clark NL, Findlay GD, Yi X, MacCoss MJ, Swanson WJ. 2007. Duplication and selection on abalone sperm lysin in an allopatric population. Molecular Biology and Evolution 24:2081–2090.

Clark NL, Gasper J, Sekino M, Springer SA, Aquadro CF, Swanson WJ. 2009. Coevolution of interacting fertilization proteins. Plos Genet 5:e1000570.

Cohen N, Wassarman PM. 2001. Association of egg zona pellucida glycoprotein mZP3 with sperm protein sp56 during fertilization in mice. Int J Dev Biol 45:569–576.

Dean MD, Clark NL, Findlay GD, Karn RC, Yi X, Swanson WJ, MacCoss MJ, Nachman MW. 2009. Proteomics and comparative genomic investigations reveal heterogeneity in evolutionary rate of male reproductive proteins in mice (Mus domesticus). Molecular Biology and Evolution 26:1733–1743.

Dorus S, Busby SA, Gerike U, Shabanowitz J, Hunt DF, Karr TL. 2006. Genomic and functional evolution of the Drosophila melanogaster sperm proteome. Nat Genet 38:1440–1445.

Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT. 2004. Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. Nat Genet 36:1326–1329.

Dorus S, Wasbrough ER, Busby J, Wilkin EC, Karr TL. 2010. Sperm proteomics reveals intensified selection on mouse sperm membrane and acrosome genes. Molecular Biology and Evolution 27:1235–1246.

Eisenbach M. 1999. Sperm chemotaxis. Rev Reprod 4:56–66.

Eng JK, McCormack AL, Yates JR III. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. Journal of the American Society for Mass Spectrometry 5:976–989.

Findlay GD, MacCoss MJ, Swanson WJ. 2009. Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in Drosophila. Genome Research 19:886–896.

Finn S, Civetta A. 2010. Sexual Selection and the Molecular Evolution of ADAM Proteins. J Mol Evol 71:231–240.

Galindo BE, Vacquier VD, Swanson WJ. 2003. Positive selection in the egg receptor for abalone sperm lysin. P Natl Acad Sci Usa 100:4639–4643.

Gavrilets S, Waxman D. 2002. Sympatric speciation by sexual conflict. P Natl Acad Sci Usa 99:10533–10538.

Gavrilets S. 2000. Rapid evolution of reproductive barriers driven by sexual conflict. Nature 403:886–889.

Geyer L, Palumbi S. 2003. Reproductive character displacement and the genetics of gamete recognition in tropical sea urchins. Evolution 57:1049–1060.

Geyer LB, Lessios HA. 2009. Lack of character displacement in the male recognition molecule, bindin, in Altantic sea urchins of the genus Echinometra. Molecular Biology and Evolution 26:2135–2146.

Glabe CG, Vacquier VD. 1978. Egg surface glycoprotein receptor for sea urchin sperm bindin. P Natl Acad Sci Usa 75:881–885.

Götz S, García-Gómez JM, Terol J, et al. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36:3420–3435.

Grabherr MG, Haas BJ, Yassour M, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652.

Guerrero A, Nishigaki T, Carneiro J, Yoshiro Tatsu, Wood CD, Darszon A. 2010. Tuning sperm chemotaxis by calcium burst timing. Dev Biol 344:52–65.

Hamm D, Mautz BS, Wolfner MF, Aquadro CF, Swanson WJ. 2007. Evidence of amino acid diversity-enhancing selection within humans and among primates at the candidate sperm-receptor gene PKDREJ. Am J Hum Genet 81:44–52.

Helinski MEH, Deewatthanawong P, Sirot LK, Wolfner MF, Harrington LC. 2012. Duration and dose-dependency of female sexual receptivity responses to seminal fluid proteins in Aedes albopictus and Ae. aegypti mosquitoes. J Insect Physiol 58:1307–1313.

Hellberg M, Moy G, Vacquier V. 2000. Positive selection and propeptide repeats promote rapid interspecific divergence of a gastropod sperm protein. Molecular Biology and Evolution 17:458–466.

Hughes J, Ward CJ, Aspinwall R, Butler R, Harris PC. 1999. Identification of a human homologue of the sea urchin receptor for egg jelly: a polycystic kidney disease-like protein. Hum Mol Genet 8:543–549.

Jiang J-Z, Zhang W, Guo Z-X, Cai C-C, Su Y-L, Wang R-X, Wang J-Y. 2011. Functional annotation of an expressed sequence tag library from Haliotis diversicolor and analysis of its plant-like sequences. Marine Genomics 4:189–196.

Jovine L, Darie CC, Litscher ES, Wassarman PM. 2005. Zona pellucida domain proteins. Annu. Rev. Biochem. 74:83–114.

Kamei N, Glabe CG. 2003. The species-specific egg receptor for sea urchin sperm adhesion is EBR1,a novel ADAMTS protein. Genes Dev 17:2502–2507.

Kaupp UB, Kashikar ND, Weyand I. 2008. Mechanisms of sperm chemotaxis. Annu Rev Physiol 70:93–117.

Kresge N, Vacquier V, Stout C. 2001. Abalone lysin: the dissolving and evolving sperm protein. Bioessays 23:95–103.

Krzywinska E, Krzywinski J. 2009. Analysis of expression in the Anopheles gambiae developing testes reveals rapidly evolving lineage-specific genes in mosquitoes. BMC Genomics 10:300.

Lee Y, OTA T, Vacquier V. 1995. Positive selection is a general phenomenon in the evolution of abalone sperm lysin. Molecular Biology and Evolution 12:231–238.

Lee Y, Vacquier V. 1992. The divergence of species-specific abalone sperm lysins is promoted by positive Darwinian selection. Biol Bull 182:97–104.

Levitan DR, Ferrell DL. 2006. Selection on gamete recognition proteins depends on sex, density, and genotype frequency. Science 312:267–269.

Levitan DR, Stapper AP. 2010. Simultaneous positive and negative frequency-dependent selection on sperm bindin, a gamete recognition protein in the sea urchin Strongylocentrotus purpuratus. Evolution 64:785–797.

Lewis CA, Leighton D, Vacquier V. 1980. Morphology of Abalone Spermatozoa before and after the Acrosome Reaction. Journal of Ultrastructure Research 72:39–46.

Lewis CA, Talbot CF, Vacquier VD. 1982. A protein from abalone sperm dissolves the egg vitelline layer by a nonenzymatic mechanism. Dev Biol 92:227–239.

Li J, Liu F, Liu X, et al. 2011. Mapping of the human testicular proteome and its relationship with that of the epididymis and spermatozoa. Mol. Cell Proteomics 10:M110.004630.

Li J, Liu F, Wang H, et al. 2010. Systematic mapping and functional analysis of a family of human epididymal secretory sperm-located proteins. Mol. Cell Proteomics 9:2517–2528.

Lyon J, Vacquier V. 1999. Interspecies chimeric sperm lysins identify regions mediating species-specific recognition of the abalone egg vitelline envelope. Dev Biol 214:151–159.

Marshall JL, Huestis DL, Garcia C, Hiromasa Y, Wheeler S, Noh S, Tomich JM, Howard DJ. 2010. Comparative proteomics uncovers the signature of natural selection acting on the ejaculate proteomes of two cricket species isolated by postmating, prezygotic phenotypes. Molecular Biology and Evolution.

Martín-Coello J, Benavent-Corai J, Roldan ERS, Gomendio M. 2009. Sperm competition promotes asymmetries in reproductive barriers between closely related species. Evolution 63:613–623.

McCartney MA, Lessios HA. 2004. Adaptive evolution of sperm bindin tracks egg

incompatibility in neotropical sea urchins of the genus Echinometra. Molecular Biology and Evolution 21:732–745.

Metz E, KANE R, YANAGIMACHI H, Palumbi S. 1994. Fertilization between closely-related sea urchins is blocked by incompatibilities during sperm-egg attachment and early stages of fusion. Biol Bull 187:23–34.

Metz EC, Palumbi SR. 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. Molecular Biology and Evolution [Internet] 13:397–406. Available from: http://apps.isiknowledge.com/InboundService.do?Func=Frame&product=WOS&action=retrieve&SrcApp=Papers&UT=A1996TU42800010&SID=2AO8n7OfjAC4n5CajMh&Init=Yes&SrcAuth=mekentosj&mode=FullRecord&customersID=mekentosj&DestFail=http%3A%2F%2Faccess.isiproducts.com%2Fcustom_images%2Fwok_failed_auth.html

Metz EC, Robles-Sikisaka R, Vacquier VD. 1998. Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. P Natl Acad Sci Usa 95:10676–10681.

Miller RL, Vogt R. 1996. An N-terminal partial sequence of the 13 kDa Pycnopodia helianthoides sperm chemoattractant "startrak" possesses sperm-attracting activity. J Exp Biol 199:311–318.

Monné M, Han L, Jovine L. 2006. Tracking down the ZP domain: From the mammalian zona pellucida to the molluscan vitelline envelope. Semin. Reprod. Med. 24:204–216.

Morita M, Nishikawa A, Nakajima A, Iguchi A, Sakai K, Takemura A, Okuno M. 2006. Eggs regulate sperm flagellar motility initiation, chemotaxis and inhibition in the coral Acropora digitifera, A. gemmifera and A. tenuis. J Exp Biol 209:4574–4579.

Morse DE, Duncan H, Hooker N, Morse A. 1977. Hydrogen peroxide induces spawning in mollusks, with activation of prostaglandin endoperoxide synthetase. Science 196:298–300.

Moy GW, Springer SA, Adams SL, Swanson WJ, Vacquier VD. 2008. Extraordinary intraspecific diversity in oyster sperm bindin. P Natl Acad Sci Usa 105:1993–1998.

Nakachi M, Moriyama H, Hoshi M, Matsumoto M. 2006. Acrosome reaction is subfamily specific in sea star fertilization. Dev Biol 298:597–604.

Nakachi M, Nakajima A, Nomura M, Yonezawa K, Ueno K, Endo T, Inaba K. 2011. Proteomic profiling reveals compartment-specific, novel functions of ascidian sperm proteins. Mol Reprod Dev 78:529–549.

Noor MA. 1999. Reinforcement and other consequences of sympatry. Heredity 83 (Pt 5):503–508.

Nydam ML, Harrison RG. 2011. Reproductive protein evolution in two cryptic species of marine chordate. BMC evolutionary biology 11:18.

Palmer MR, McDowall MH, Stewart L, Ouaddi A, MacCoss MJ, Swanson WJ. 2013. Mass spectrometry and next-generation sequencing reveal an abundant and rapidly evolving abalone sperm protein. Mol Reprod Dev 80:460–465.

Palumbi S. 1999. All males are not created equal: Fertility differences depend on gamete recognition polymorphisms in sea urchins. P Natl Acad Sci Usa 96:12632–12637.

Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, Conaway RC, Conaway JW, Florens L, Washburn MP. 2006. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. P Natl Acad Sci Usa 103:18928–18933.

Park S-C, Kim J-Y, Lee J-K, Hahm K-S, Park Y. 2012. Antibacterial Action of New Antibacterial Peptides, Nod1 and Nod2, Isolated from Nordotis discus discus. J Agric Food Chem.

Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, Lehrach H, Soldatov A. 2009. Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res 37:e123.

Petersen TN, Brunak S, Heijne von G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786.

Pujolar JM, Pogson GH. 2011. Positive Darwinian selection in gamete recognition proteins of Strongylocentrotus sea urchins. Mol Ecol 20:4968–4982.

Ramm SA, McDonald L, Hurst JL, Beynon RJ, Stockley P. 2009. Comparative proteomics reveals evidence for evolutionary diversification of rodent seminal fluid and its functional significance in sperm competition. Molecular Biology and Evolution 26:189–198.

Riffell JA, Krug PJ, Zimmer RK. 2004. The ecological and evolutionary consequences of sperm chemoattraction. P Natl Acad Sci Usa 101:4501–4506.

Riginos C, Wang D, Abrams AJ. 2006. Geographic variation and positive selection on M7 lysin, an acrosomal sperm protein in mussels (Mytilus spp.). Molecular Biology and Evolution 23:1952–1965.

Rohlfs RV, Swanson WJ, Weir BS. 2010. Detecting coevolution through allelic association between physically unlinked loci. Am J Hum Genet 86:674–685.

Rubinstein E, Ziyyat A, Wolf J-P, Le Naour F, Boucheix C. 2006. The molecular players of sperm-egg fusion in mammals. Semin Cell Dev Biol 17:254–263.

Schully SD, Hellberg ME. 2006. Positive selection on nucleotide substitutions and indels

in accessory gland proteins of the Drosophila pseudoobscura subgroup. J Mol Evol 62:793–802.

Sharma V, Eng JK, MacCoss MJ, Riffle M. 2012. A mass spectrometry proteomics data management platform. Mol. Cell Proteomics 11:824–831.

Short SM, Wolfner MF, Lazzaro BP. 2012. Female Drosophila melanogaster suffer reduced defense against infection due to seminal fluid components. J Insect Physiol 58:1192–1201.

Sirot LK, Hardstone MC, Helinski MEH, Ribeiro JMC, Kimura M, Deewatthanawong P, Wolfner MF, Harrington LC. 2011. Towards a semen proteome of the dengue vector mosquito: protein identification and potential functions. PLoS Negl Trop Dis 5:e989.

Slaughter C, McCartney MA, Yund PO. 2008. Comparison of gamete compatibility between two blue mussel species in sympatry and in allopatry. Biol Bull 214:57–66.

Spehr M, Gisselmann G, Poplawski A, Riffell JA, Wetzel CH, Zimmer RK, Hatt H. 2003. Identification of a testicular odorant receptor mediating human sperm chemotaxis. Science 299:2054–2058.

Springer SA, Crespi BJ. 2007. Adaptive gamete-recognition divergence in a hybridizing Mytilus population. Evolution 61:772–783.

Stein K, Go J, Lane W, Primakoff P, Myles D. 2006. Proteomic analysis of sperm regions that mediate sperm-egg interactions. Proteomics 6:3533–3543.

Sun F, Giojalas LC, Rovasio RA, Tur-Kaspa I, Sanchez R, Eisenbach M. 2003. Lack of species-specificity in mammalian sperm chemotaxis. Dev Biol 255:423–427.

Sutton KA, Jungnickel MK, Florman HM. 2008. A polycystin-1 controls postcopulatory reproductive selection in mice. P Natl Acad Sci Usa 105:8661–8666.

Swanson W, Vacquier V. 1995a. Extraordinary divergence and positive Darwinian selection in a fusagenic protein coating the acrosomal process of abalone spermatozoa. P Natl Acad Sci Usa 92:4957–4961.

Swanson W, Vacquier V. 1995b. Liposome fusion induced by a M(R)-18000 protein localized to the acrosomal region of acrosome-reacted abalone spermatozoa. Biochemistry-Us 34:14202–14208.

Swanson W, Vacquier V. 1997. The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule. P Natl Acad Sci Usa 94:6724–6729.

Swanson W, Vacquier V. 1998. Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. Science 281:710–712.

Swanson W, Vacquier V. 2002. Reproductive protein evolution. Annu Rev Ecol Syst

33:161–179.

Swanson WJ, Aagaard JE, Vacquier VD, Monné M, Sadat Al Hosseini H, Jovine L. 2011. The molecular basis of sex: linking yeast to human. Molecular Biology and Evolution 28:1963–1966.

Swanson WJ, Nielsen R, Yang Q. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. Molecular Biology and Evolution 20:18–20.

Swanson WJ, Yang Z, Wolfner MF, Aquadro CF. 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. P Natl Acad Sci Usa 98:2509–2514.

Tal M, Silberstein A, Nusser E. 1985. Why does Coomassie Brilliant Blue R interact differently with different proteins? A partial answer. J Biol Chem 260:9976–9980.

Tomczuk M, Takahashi Y, Huang J, et al. 2003. Role of multiple beta1 integrins in cell adhesion to the disintegrin domains of ADAMs 2 and 3. Exp Cell Res 290:68–81.

Turner LM, Hoekstra HE. 2006. Adaptive evolution of fertilization proteins within a genus: variation in ZP2 and ZP3 in deer mice (Peromyscus). Molecular Biology and Evolution 23:1656–1669.

Ulrich AS, Otter M, Glabe CG, Hoekstra D. 1998. Membrane fusion is induced by a distinct peptide sequence of the sea urchin fertilization protein bindin. J Biol Chem 273:16748–16755.

Vacquier VD, Moy GW. 1977. Isolation of bindin: the protein responsible for adhesion of sperm to sea urchin eggs. P Natl Acad Sci Usa 74:2456–2460.

Vacquier VD, Moy GW. 1997. The fucose sulfate polymer of egg jelly binds to sperm REJ and is the inducer of the sea urchin sperm acrosome reaction. Dev Biol 192:125–135.

Vilela-Silva A-CES, Hirohashi N, Mourao PAS. 2008. The structure of sulfated polysaccharides ensures a carbohydrate-based mechanism for species recognition during sea urchin fertilization. Int J Dev Biol 52:551–559.

Wagstaff BJ, Begun DJ. 2005. Molecular population genetics of accessory gland protein genes and testis-expressed genes in Drosophila mojavensis and D. arizonae. Genetics 171:1083–1101.

Walters JR, Harrison RG. 2010. Combined EST and proteomic analysis identifies rapidly evolving seminal fluid proteins in Heliconius butterflies. Molecular Biology and Evolution 27:2000–2013.

Ward GE, Brokaw CJ, Garbers DL, Vacquier VD. 1985. Chemotaxis of Arbacia punctulata spermatozoa to resact, a peptide from the egg jelly layer. J Cell Biol

101:2324–2329.

Wasbrough ER, Dorus S, Hester S, Howard-Murkin J, Lilley K, Wilkin E, Polpitiya A, Petritis K, Karr TL. 2010. The Drosophila melanogaster sperm proteome-II (DmSP-II). J Proteomics 73:2171–2185.

Wolfner MF. 2009. Battle and ballet: molecular interactions between the sexes in Drosophila. J. Hered. 100:399–410.

Wong JL, Johnson MA. 2010. Is HAP2-GCS1 an ancestral gamete fusogen? Trends Cell Biol 20:134–141.

Yamada L, Saito T, Taniguchi H, Sawada H, Harada Y. 2009. Comprehensive Egg Coat Proteome of the Ascidian Ciona intestinalis Reveals Gamete Recognition Molecules Involved in Self-sterility. J Biol Chem 284:9402–9410.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13:555–556.

Yauger B, Boggs NA, Dean J. 2011. Human ZP4 is not sufficient for taxon-specific sperm recognition of the zona pellucida in transgenic mice. Reproduction 141:313–319.

Zasloff M. 2002. Antimicrobial peptides of multicellular organisms. Nature 415:389–395.

Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP. 2006. Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. J. Proteome Res. 5:2339–2347.

Appendix A

Supplemental Materials for Chapter 2:

Other Supplementary Materials for this chapter includes the following:

- DNA and protein sequences for genes analyzed with codeml are available in GenBank, accession numbers KC752594 – KC752618.
- RNA-seq reads are available in the Short Read Archive, accession number SRR770266.
- Raw mass spectrometry data files and the transcriptome assembly used for the mass spectrometry database are available for download on the lab website (http://depts.washington.edu/swansonw/Swanson_Lab/Data.html).

Supplementary Text

Transcriptome and proteome

The lengths range from 200 to 4,898 bp with an average of 472 bp. In a BLAST search to the NCBI non-redundant protein database, 6,735 of these have a significant match, with an e-value cutoff of 10^{-3}.

The Triton-extracted whole sperm sample (one biological replicate) contains 490 protein groups, the acrosome-enriched samples (three biological replicates) contain 480 protein groups, and the extracellular digestion samples (three biological replicates) contain 506 protein groups. Protein groups represent either a single protein with redundant RNA-seq contigs, or a protein with multiple forms. 189 proteins are shared between all three preparations.

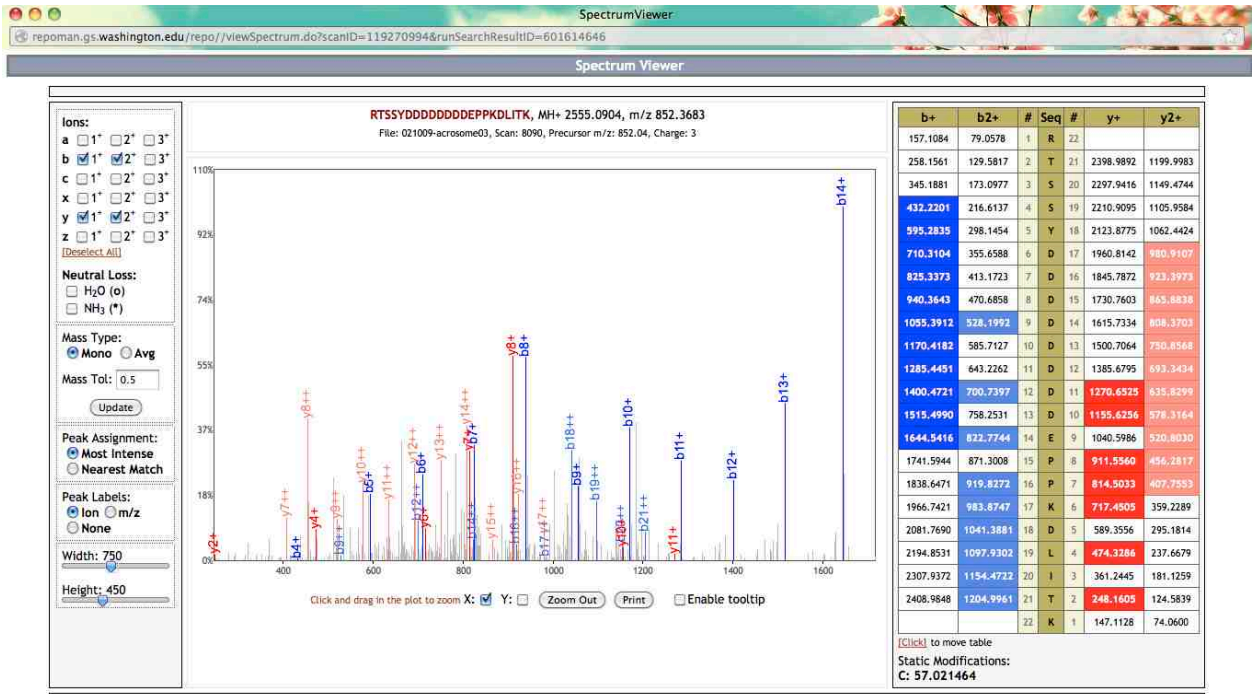
BLAST results

Candidate genes were chosen based on abundance, with particular interest in proteins predicted to be secreted, transmembrane, abundant, or without a close BLAST match from which to infer function. Of 883 proteins, 57 have a predicted signal sequence, and of the 676 BLAST match top-hit proteins, 37 have a signal sequence, as predicted by SignalP 4.0 (Petersen et al. 2011). 183 of the abalone proteins have at least one predicted transmembrane helix, as predicted by HMM (TMHMM Server v. 2.0, CBS Technical University of Denmark).

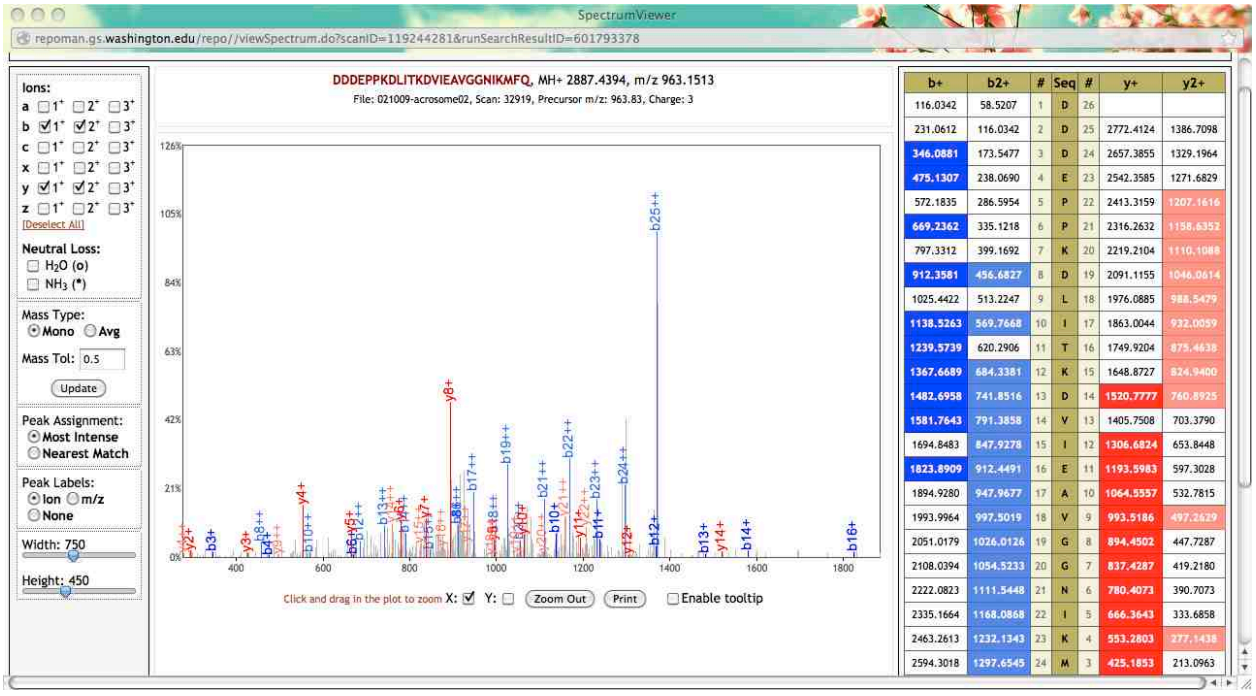
Supplementary Figures

Figure S1. Representative spectra for coverage of sp6 proteins from tandem mass spectrometry for sp6_8D (A) and sp6_4D (B).

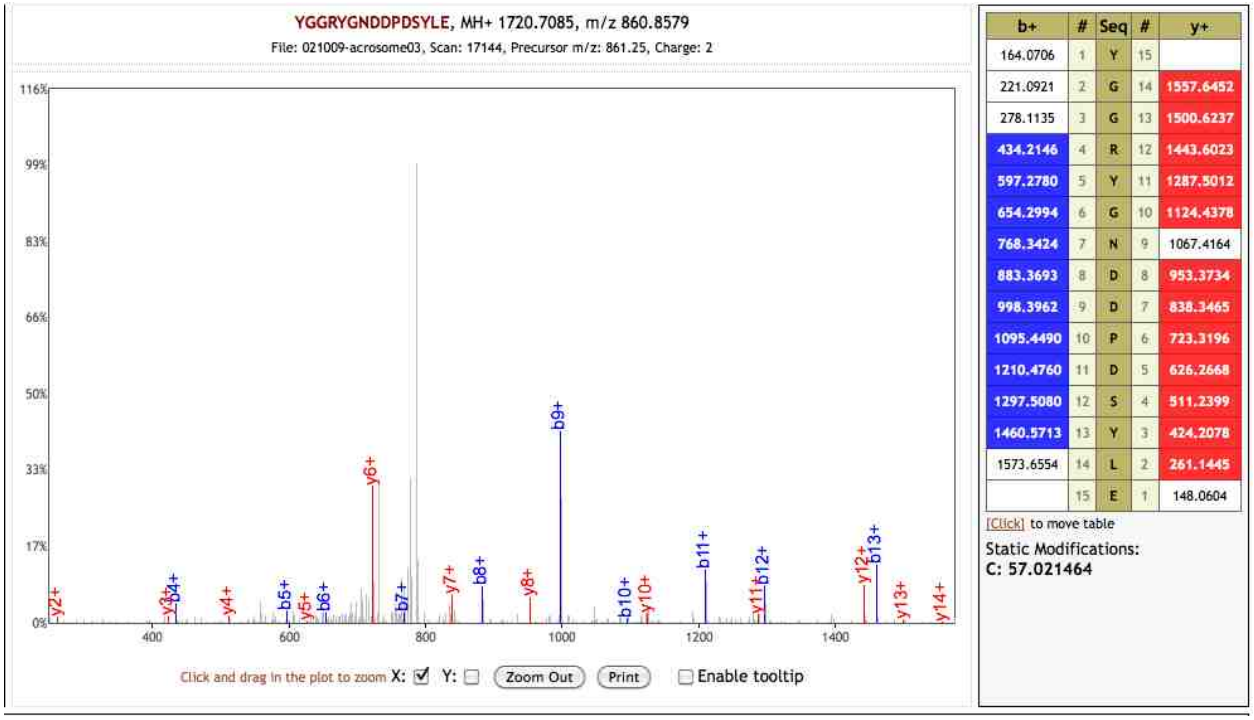
A) sp6_8D



b+	b2+	#	Seq	#	y+	y2+
157.1084	79.0578	1	R	22		
258.1561	129.5817	2	T	21	2398.9892	1199.9983
345.1881	173.0977	3	S	20	2297.9416	1149.4744
432.2201	216.6137	4	S	19	2210.9095	1105.9584
595.2835	298.1454	5	Y	18	2123.8775	1062.4424
710.3104	355.6588	6	D	17	1960.8142	980.9107
825.3373	413.1723	7	D	16	1845.7872	923.3973
940.3643	470.6858	8	D	15	1730.7603	865.8838
1055.3912	528.1992	9	D	14	1615.7334	808.3703
1170.4182	585.7127	10	D	13	1500.7064	750.8568
1285.4451	643.2262	11	D	12	1385.6795	693.3434
1400.4721	700.7397	12	D	11	1270.6525	635.8299
1515.4990	758.2531	13	D	10	1155.6256	578.3164
1644.5416	822.7744	14	E	9	1040.5986	520.8030
1741.5944	871.3008	15	P	8	911.5560	456.2817
1838.6471	919.8272	16	P	7	814.5033	407.7553
1966.7421	983.8747	17	K	6	717.4505	359.2289
2081.7690	1041.3881	18	D	5	589.3556	295.1814
2194.8531	1097.9302	19	L	4	474.3286	237.6679
2307.9372	1154.4722	20	I	3	361.2445	181.1259
2408.9848	1204.9961	21	T	2	248.1605	124.5839
		22	K	1	147.1128	74.0600



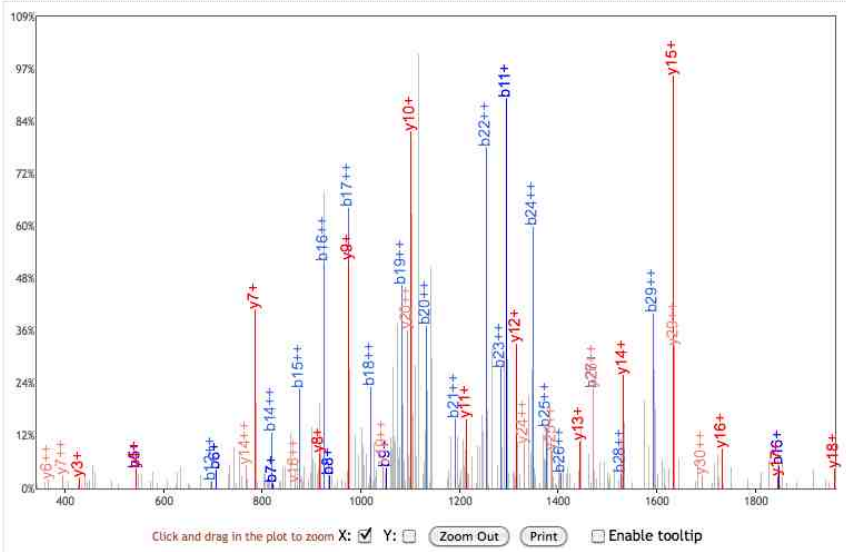
b+	b2+	#	Seq	#	y+	y2+
116.0342	58.5207	1	D	26		
231.0612	116.0342	2	D	25	2772.4124	1386.7098
346.0881	173.5477	3	D	24	2657.3855	1329.1964
475.1307	238.0690	4	E	23	2542.3585	1271.6829
572.1835	286.5954	5	P	22	2413.3159	1207.1616
669.2362	335.1218	6	P	21	2316.2632	1158.6352
797.3312	399.1692	7	K	20	2219.2104	1110.1088
912.3581	456.6827	8	D	19	2091.1155	1046.0614
1025.4422	513.2247	9	L	18	1976.0885	988.5479
1138.5263	569.7668	10	I	17	1863.0044	932.0059
1239.5739	620.2906	11	T	16	1749.9204	875.4638
1367.6689	684.3381	12	K	15	1648.8727	824.9400
1482.6958	741.8516	13	D	14	1520.7777	760.8925
1581.7643	791.3858	14	V	13	1405.7508	703.3790
1694.8483	847.9278	15	I	12	1306.6824	653.8448
1823.8909	912.4491	16	E	11	1193.5983	597.3028
1894.9280	947.9677	17	A	10	1064.5557	532.7815
1993.9964	997.5019	18	V	9	993.5186	497.2629
2051.0179	1026.0126	19	G	8	894.4502	447.7287
2108.0394	1054.5233	20	G	7	837.4287	419.2180
2222.0823	1111.5448	21	N	6	780.4073	390.7073
2335.1664	1168.0868	22	I	5	666.3643	333.6858
2463.2613	1232.1343	23	K	4	553.2803	277.1438
2594.3018	1297.6545	24	M	3	425.1853	213.0963


YGGRYGNDDPDSYLE, MH+ 1720.7085, m/z 860.8579
File: 021009-acrosome03, Scan: 17144, Precursor m/z: 861.25, Charge: 2
116%
99%
83%
66%
50%
33%
17%
0%
400
600
800
1000
1200
1400
y2+
y3+
b4+
y4+
b5+
y5+
b6+
y6+
b7+
y7+
b8+
y8+
b9+
b10+
y10+
b11+
y11
b12+
y12+
b13+
y13+
y14+
Click and drag in the plot to zoom X: Y: Zoom Out Print Enable tooltip
b+ # Seq # y+
164.0706 1 Y 15
221.0921 2 G 14 1557.6452
278.1135 3 G 13 1500.6237
434.2146 4 R 12 1443.6023
597.2780 5 Y 11 1287.5012
654.2994 6 G 10 1124.4378
768.3424 7 N 9 1067.4164
883.3693 8 D 8 953.3734
998.3962 9 D 7 838.3465
1095.4490 10 P 6 723.3196
1210.4760 11 D 5 626.2668
1297.5080 12 S 4 511.2399
1460.5713 13 Y 3 424.2078
1573.6554 14 L 2 261.1445
15 E 1 148.0604
[Click] to move table
Static Modifications:
C: 57.021464

B) sp6_4D

VLSSRYDDDDEPEDLVTSETIQGMGGEDMYD, MH+ 3481.4417, m/z 1161.1521

File: 021009-acrosome02, Scan: 27382, Precursor m/z: 1161.81, Charge: 3



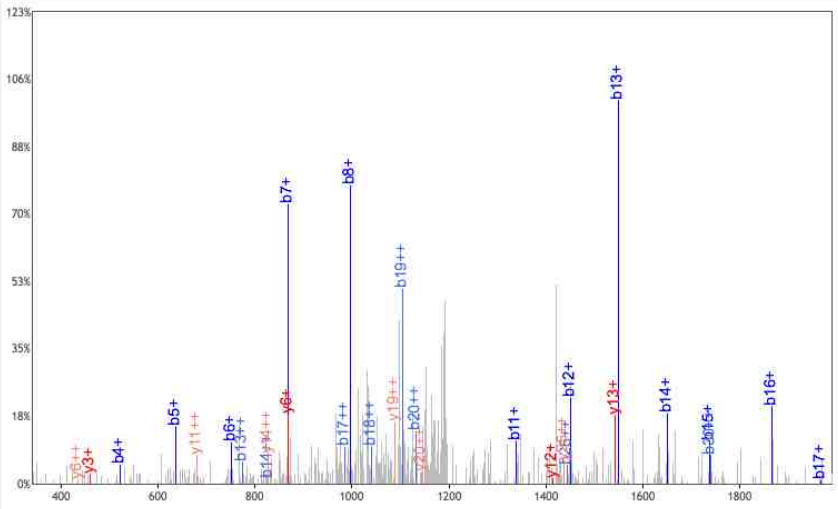
b+	b2+	#	Seq	#	y+	y2+
100.0757	50.5415	1	V	31		
213.1598	107.0835	2	L	30	3382.3733	1691.6903
300.1918	150.5995	3	S	29	3269.2893	1635.1483
387.2238	194.1155	4	S	28	3182.2572	1591.6323
543.3249	272.1661	5	R	27	3095.2252	1548.1162
706.3883	353.6978	6	Y	26	2939.1241	1470.0657
821.4152	411.2112	7	D	25	2776.0608	1388.5340
936.4421	468.7247	8	D	24	2661.0338	1331.0206
1051.4691	526.2382	9	D	23	2546.0069	1273.5071
1166.4960	583.7516	10	D	22	2430.9799	1215.9936
1295.5386	648.2729	11	E	21	2315.9530	1158.4801
1392.5914	696.7993	12	P	20	2186.9104	1093.9588
1521.6340	761.3206	13	E	19	2089.8576	1045.4325
1636.6609	818.8341	14	D	18	1960.8151	980.9112
1749.7450	875.3761	15	L	17	1845.7881	923.3977
1848.8134	924.9103	16	V	16	1732.7040	866.8557
1949.8611	975.4342	17	T	15	1633.6356	817.3215
2036.8931	1018.9502	18	S	14	1532.5880	766.7976
2165.9357	1083.4715	19	E	13	1445.5559	723.2816
2266.9834	1133.9953	20	T	12	1316.5133	658.7603
2380.0674	1190.5374	21	I	11	1215.4657	608.2365
2508.1260	1254.5666	22	Q	10	1102.3816	551.6944
2565.1475	1283.0774	23	G	9	974.3230	487.6651
2696.1880	1348.5976	24	M	8	917.3015	459.1544
2753.2094	1377.1084	25	G	7	786.2611	393.6342
2810.2309	1405.6191	26	G	6	729.2396	365.1234
2939.2735	1470.1404	27	E	5	672.2181	336.6127
3054.3004	1527.6539	28	D	4	543.1755	272.0914
3185.3409	1593.1741	29	M	3	428.1486	214.5779
3348.4042	1674.7058	30	Y	2	297.1081	149.0577
		31	D	1	134.0448	67.5260

[Click] to move table

Static Modifications:
C: 57.021464

SRYDDDDEPEDLVTSETIQGMGGEDMYDRRK, MH+ 3622.5544, m/z 1208.1897

File: 110909_elast_1, Scan: 27610, Precursor m/z: 1208.19, Charge: 3



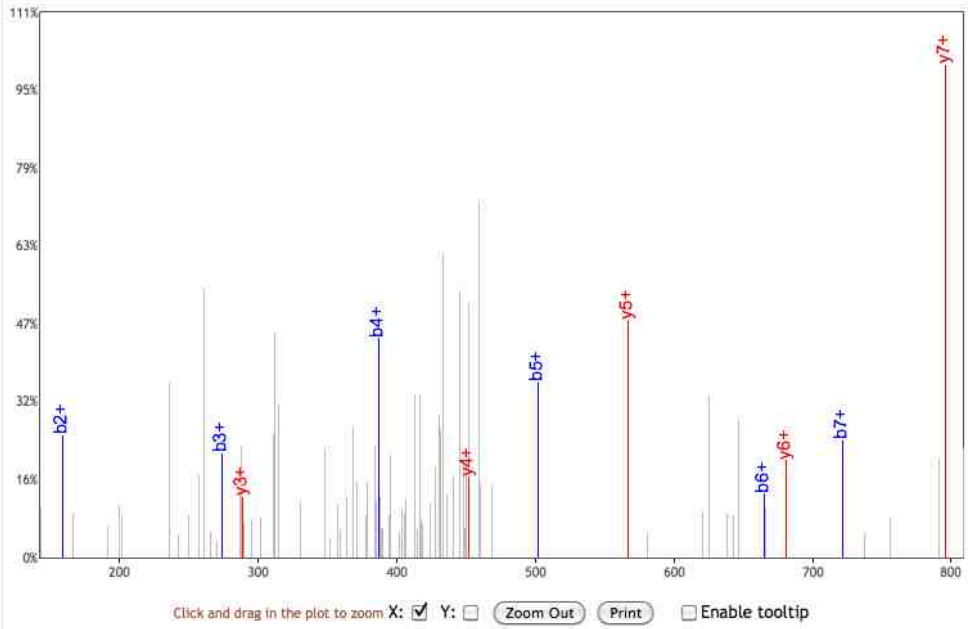
b+	b2+	#	Seq	#	y+	y2+
88.0393	44.5233	1	S	31		
244.1404	122.5738	2	R	30	3535.5224	1768.2648
407.2037	204.1055	3	Y	29	3379.4213	1690.2143
522.2307	261.6190	4	D	28	3216.3580	1608.6826
637.2576	319.1325	5	D	27	3101.3310	1551.1691
752.2846	376.6459	6	D	26	2986.3041	1493.6557
867.3115	434.1594	7	D	25	2871.2771	1436.1422
996.3541	498.6807	8	E	24	2756.2502	1378.6287
1093.4069	547.2071	9	P	23	2627.2076	1314.1074
1222.4495	611.7284	10	E	22	2530.1548	1265.5811
1337.4764	669.2418	11	D	21	2401.1122	1201.0598
1450.5605	725.7839	12	L	20	2286.0853	1143.5463
1549.6289	775.3181	13	V	19	2173.0012	1087.0043
1650.6766	825.8419	14	T	18	2073.9328	1037.4700
1737.7086	869.3579	15	S	17	1972.8851	986.9462
1866.7512	933.8792	16	E	16	1885.8531	943.4302
1967.7989	984.4031	17	T	15	1756.8105	878.9089
2080.8829	1040.9451	18	I	14	1655.7628	828.3851
2208.9415	1104.9744	19	Q	13	1542.6788	771.8430
2265.9630	1133.4851	20	G	12	1414.6202	707.8137
2397.0035	1199.0054	21	M	11	1357.5987	679.3030
2454.0249	1227.5161	22	G	10	1226.5582	613.7828
2511.0464	1256.0268	23	G	9	1169.5368	585.2720
2640.0890	1320.5481	24	E	8	1112.5153	556.7613
2755.1159	1378.0616	25	D	7	983.4727	492.2400
2886.1564	1443.5818	26	M	6	868.4458	434.7265
3049.2197	1525.1135	27	Y	5	737.4053	369.2063
3164.2467	1582.6270	28	D	4	574.3420	287.6746
3320.3478	1660.6775	29	R	3	459.3150	230.1612
3476.4489	1738.7281	30	R	2	303.2139	152.1106
		31	K	1	147.1128	74.0600

[Click] to move table

Static Modifications:
C: 57.021464

SADLDYGGR, MH+ 953.4323, m/z 477.2198

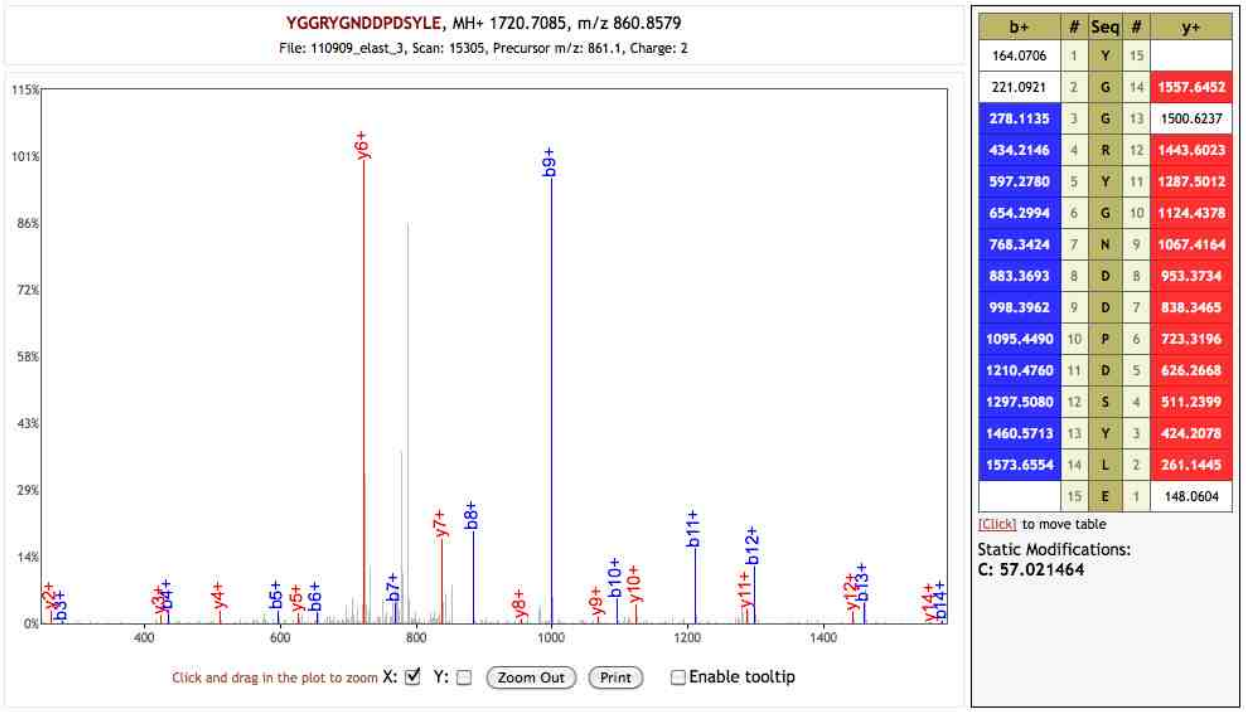
File: 110909_tryp_2, Scan: 5380, Precursor m/z: 477.37, Charge: 2



b+	#	Seq	#	y+
88.0393	1	S	9	
159.0764	2	A	8	866.4003
274.1034	3	D	7	795.3632
387.1874	4	L	6	680.3362
502.2144	5	D	5	567.2522
665.2777	6	Y	4	452.2252
722.2992	7	G	3	289.1619
779.3206	8	G	2	232.1404
	9	R	1	175.1190

[Click] to move table

Static Modifications:
C: 57.021464



b+	#	Seq	#	y+
164.0706	1	Y	15	
221.0921	2	G	14	1557.6452
278.1135	3	G	13	1500.6237
434.2146	4	R	12	1443.6023
597.2780	5	Y	11	1287.5012
654.2994	6	G	10	1124.4378
768.3424	7	N	9	1067.4164
883.3693	8	D	8	953.3734
998.3962	9	D	7	838.3465
1095.4490	10	P	6	723.3196
1210.4760	11	D	5	626.2668
1297.5080	12	S	4	511.2399
1460.5713	13	Y	3	424.2078
1573.6554	14	L	2	261.1445
	15	E	1	148.0604

Figure S2. sp6 alignments used to calculate d_N/d_S, result in Table 1.

```
                                25        30        35        40        45        50        55        60        65        70        75
H. rufescens 4D                 LSSRYDDDDEPEDLVTSETIQGMGGED-MYDRRKRSADLDYGGRYGNDDPDSYLE*--
H. kamtschatkana 4Ds            SSSAYDDDDEPDDLIRREVTEGVGGDEEMYKRRKRSADLDYGGGYGNDDPDNYLESK*
H. fulgens 4D                   SGGAFDDDDTPENLVSHKVIDARA-EEELG-RRRRSADLDYGGGYADSDGSESLLE*-
```

```
                                25        30        35        40        45        50        55        60        65        70        75        80        85
H. discus                       YIDDDDDDDDDDDEPGDLVPGVVTPAVDGDEGPFGRKKRAVELDYGAGYGDESDDSLLN*--
H. rufescens 8D                 YDDDDDDDDEP---PKDLITKDVIEAVGGNIKMFQRRKRSADLDYGGRYGNDDPDSYLE---
H. kamtschatkana 2D             YDDEDEDD--SDEEPKDLITNEVVEAVGGSEAMFQRRKRSADLDYGGGYGNDDPDNYLESK*
H. kamtschatkana 4Dl            YDDDDEDEDDSDEEPKDLITKEVVEAMGGSESMFQRRKRSADSDYGGGYGNDDPDNYLESK*
```

Supplementary Methods

Transcriptome Sequencing

RNA was extracted from the testes of 5 small *H. rufescens* by cesium chloride gradient. mRNA was isolated with the Qiagen Oligotex mRNA mini kit, standard protocol, followed by DNase treatment. With approximately 1µg of mRNA, strand specific cDNA was synthesized as in (Parkhomchuk et al. 2009). The cDNA was fragmented, end-repaired, a-tailed, ligated to adapters, and amplified with BioRad iProof High Fidelity polymerase. The products were size selected for 200-400bp. 76 base pair

paired-end reads were generated on an Illumina Genome Analyzer. Illumina paired-end sequencing of *H. rufescens* testis cDNA yielded 39,673,235 76 bp read pairs. Quality filtering with a phred score cutoff of 5 pared that down to a total of 19,746,048 pairs of reads and 8,661,470 single reads for a total of 48,153,556 reads. 1,300 contigs were added from Sanger sequencing of a testis cDNA library. Assembly was performed with the Trinity package (Grabherr et al. 2011). After a 200 bp cutoff, 41,576 contigs remain in the assembly. Read coverage was estimated with the BWA and Samtools packages.

Proteomics

Spawning was induced in *H. rufescens* males in 0.2µm filtered seawater with the Tris/H_2O_2 protocol (Morse et al. 1977). Intact sperm were treated with trypsin or elastase (Promega) to extract extracellular proteins, or induced to acrosome react by addition of a calcium ionophore (Sigma A23187) to 0.2 mg/ml and incubation on ice until most sperm were acrosome reacted in a subsample viewed with a microscope, or spun down and homogenized intact. All protein preparations were digested with trypsin or elastase. A 40cm fused silica 75µm internal diameter column was packed with 90Å Jupiter resin for in-line HPLC, and peptides were eluted over a four-hour gradient through the column and into an LTQ. Mass spectra were searched against a six-frame translation of the transcriptome assembly for protein identification, with a decoy database of reversed sequences, using Sequest (Eng et al. 1994). Data were visualized and summarized with the University of Washington's Yeast Resource Center's Mass Spectrometry Data Platform: MSDaPl (http://repoman.gs.washington.edu/repo/pages/internal/front.jsp). Relative abundance of proteins in a sample was estimated by the normalized spectral abundance factor (NSAF) (Paoletti et al. 2006).Two or three technical replicates were performed per sample, with three spawning events for biological replicates. Sperm samples were pooled from different males with each spawning event.

Western blot and immunofluorescence

Sp6 was purified with anion-exchange chromatography and confirmed with SDS-PAGE. Polyclonal antibodies were raised against purified sp6 from *H. rufescens* testis in rabbits by R&R Research, LLC (Stanwood, WA), and diluted serum was used for Western blots and sperm immunofluorescence.

Proteins were separated on 15% SDS-PAGE gels, and proteins transferred to Hybond™ -P PVDF transfer membrane with electrophoresis. Membranes were blocked with 3% non-fat dry milk in TBS for 1 hour at room temperature, incubated with various concentrations of anti-sera, and with peroxidase conjugated goat anti-rabbit IgG (Pierce) at 1:2,500 for 1 hour each at room temperature. The signal was detected with ECL Plus Western Blotting Detection Reagents, and exposed to Hyperfilm™ ECL high performance chemiluminescence film (Amersham), developed in a Kodak X-ray developer.

For immunofluorescence, sperm from testis were fixed in 3% paraformaldehyde, 0.1% gluteraldehyde for 1 hour at room temperature. Sperm were sedimented by centrifugation at 3,000 x g for 10 minutes, and resuspended in TBS, 100mM glycine, 10mM sodium azide 5 times to wash. To permeabilize the membranes, sperm were resuspended and incubated in 100% methanol for 15 minutes with shaking. Sperm were blocked in 1% BSA in TBS, 10mM sodium azide, over night at 4°C, incubated in 1:25

sp6 anti-serum or pre-injection serum in blocking buffer for 1 hour at room temperature, washed 5 times in blocking buffer, then incubated in 1:100 rhodamine conjugated goat anti-rabbit IgG (Thermo #31686) in the dark, washed five more times, and resuspended in TBS, 0.2% Triton X-100, 10mM sodium azide. Sperm were visualized immediately on a compound light and fluorescence microscope.

Sequence analysis

To determine homology to known proteins, Blast2GO (Götz et al. 2008) was run against the non-human, non-mouse EST database, and the non-redundant protein database in NCBI. The protein and transcript sequences were also searched for transmembrane domains and signal sequences, using HMM (THMM Server v. 2.0, CGS Technical University of Denmark) and SignalP (Petersen et al. 2011). Protein molecular weights and charges were estimated with tools on the ExPASy Bioinformatics Resource Portal. From *H. rufescens* candidate sequences, primer sequences were designed to PCR amplify the transcripts from testis cDNA of other species. d_N/d_S analyses were done with the codeml package in PAML (Yang 1997).

Primers and adapters

Adapter sequences for Illumina sequencing

1. /5phos/GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
2. ACACTCTTTCCCTACACGACGCTCTTCCGATC*T. ("*" is a phosphorothioate bond)

Primer sequences for Illumina sequencing:

F: 5’
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC
GATCT
R: 5’
CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTC
TTCCGATCT

PCR Primers:

sp6_forward:
5’GGGCGTTTTAGACTGTTCAACTATCGCACTCAAC3’
sp6_reverse:
5’CTGGCCTCCTTGTGCTGACTATTTCCCGGTCAATC3’

Appendix B

Supplementary Materials for Chapter 3

Expression constructs

V5 tagged construct sequences, synthesized in pIDTSMART-KAN plasmids from IDT:

>sp6_V5_4Dx

GCG CGC GCC ATGggcaaaccgattccgaacccgctgctgggcctggatagcacc GGA GTT TTA AGC AGC CGT TAT GAC GAT GAC GAC GAA CCG GAA GAC CTG GTT ACC AGT GAG ACT ATT CAG GGT ATG GGC GGC GAA GAT ATG TAT GAC CGT CGT AAA CGC TCC GCA GAT CTG GAC TAT GGT GGC CGT TAC GGG AAT GAT GAT CCC GAT TCA TAT CTG GAG TGA TGA TAA GGA TCC CGC GCG

Amino acid sequence:

Expected size:

>sp6_V5_8Dx

GCG CGC GCC ATGggcaaaccgattccgaacccgctgctgggcctggatagcacc GGA CGC ACT TCA TCG TAT GAT GAC GAC GAT GAC GAT GAT GAC GAA CCA CCC AAA GAC CTC ATC ACG AAA GAT GTG ATT GAA GCC GTC GGC GGC AAT ATT AAA ATG TTT CAA CGT CGT AAA CGT AGC GCT GAC TTA GAT TAT GGG GGC CGC TAT GGG AAT GAT GAT CCG GAT AGT TAC TTA GAA TGA TGA TAA GGA TCC CGC GCG

Amino acid sequence:

Expected size:

>sp6_V5_scrx

gcg cgc gcc atg ggcaaaccgattccgaacccgctgctgggcctggatagcacc gga TCG GCG CCA CTG CTG AGC GAT GAT GAA GTC GAT CGC GGC CGT CTG AGC GTG ACG TCC GGG GGT ATT TCA GAT GAA CCG CAG TAT GAT ACC GAC GAC CGT GAG CGC AAC GAT GGC GAT ATG GGC GAA GAT GGC AAG TAT AGC TAT GAA CTG TAC CGT ATG GAT TAC GAC tga taa gga tcc cgc gcg

Amino acid sequence:

Expected size:

>sp6_V5_4Dx

MGKPIPNPLLGLDSTGVLSSRYDDDDEPEDLVTSETIQGMGGEDMYDRRKRSAD LDYGGRYGNDDPDSYLE

Molecular Weight: 7875.44

pI: 3.98

>sp6_V5_8Dx

MGKPIPNPLLGLDSTGRTSSYDDDDDDDDEPPKDLITKDVIEAVGGNIKMFQRRK RSADLDYGGRYGNDDPDSYLE

Molecular Weight: 8451.13

pI: 4.18

>sp6_V5_scrx

MGKPIPNPLLGLDSTGSAPLLSDDEVDRGRLSVTSGGISDEPQYDTDDRERNDGD
MGEDGKYSYELYRMDYD
Molecular Weight: 7962.52
pI: 3.98
>sp6_4Dx
MGVLSSRYDDDDEPEDLVTSETIQGMGGEDMYDRRKRSADLDYGGRYGNDDP
DSYLE
Molecular Weight: 6471.80
pI: 3.93
>sp6_8Dx
MGRTSSYDDDDDDDDEPPKDLITKDVIEAVGGNIKMFQRRKRSADLDYGGRYG
NDDPDSYLE
Molecular Weight: 7047.49
pI: 4.13
>sp6_Scrx
MGSAPLLSDDEVDRGRLSVTSGGISDEPQYDTDDRERNDGDMGEDGKYSYELY
RMDYD
Molecular Weight: 6558.88
pI: 3.93

Table 1. Out of the top 50 in NSAF rank from all affinity column mass spec samples, these were unique to sp6 columns. OspG: Ovary proteins, sp6 as ligand, glycine eluate. OspN: Ovary proteins, sp6 as ligand, NaCl eluate. VspE: VE proteins, sp6 as ligand, glycine eluate. The numbers in the sample names refer to the date in month/day/year format.

Sample	NSAF Rank	FastaID	In what samples?	Number Peptides	Number Spectra
OspG_030511	10	1_HalRu_15_H3.g_H03_15.ab1_2	OspG1, OspG2	2	22
OspG_090811	28	1_HalRu_15_H3.g_H03_15.ab1_2	OspG1, OspG2	1	12
OspG_090811	29	HalRu_40_Testis_F12.g_F12.ab1_6	OspG2, VspE	6	15
VspE_030511	18	HalRu_40_Testis_F12.g_F12.ab1_6	OspG2, VspE	4	24
OspN_090811	28	1_HalRu_49_G02_HalRu_49_G2.b_14.ab1_4	OspN2, VspE	3	18
VspE_030511	4	1_HalRu_49_G02_HalRu_49_G2.b_14.ab1_4	OspN2, VspE	2	44
OspN_090811	26	1_HalRu_6_H1.g_H01_15.ab1_2	OspN2, VspE	1	19
VspE_030511	11	1_HalRu_6_H1.g_H01_15.ab1_2	OspN2, VspE	1	21
OspN_090811	13	1_HalRu_10_D1.g_D01_07.ab1_6	unique	1	27
VspE_030511	7	1_HalRu_10_H12.g_H12_16.ab1_4	unique	5	34
OspN_090811	12	1_HalRu_11_E6.g_E06_10.ab1_6	unique	3	18

OspN_090811	44	1_HalRu_12_F1.g_F01_11.ab1_1	unique	3	8
OspN_090811	22	1_HalRu_13_F6.g_F06_12.ab1_4	unique	1	24
OspN_090811	43	1_HalRu_18_A6.g_A06_02.ab1_2	unique	2	16
OspG_090811	11	1_HalRu_18_G11.g_G11_13.ab1_3	unique	3	18
OspG_030511	21	1_HalRu_18_G12.g_G12_14.ab1_2	unique	1	19
OspN_090811	40	1_HalRu_19_D5.g_D05_07.ab1_1	unique	2	16
OspG_030511	5	1_HalRu_20_E10.g_E10_10.ab1_1	unique	1	26
OspN_090811	33	1_HalRu_21_D9.g_D09_07.ab1_1	unique	1	20
OspG_030511	16	1_HalRu_22_D3.g_D03_07.ab1_2	unique	1	19
OspN_090811	7	1_HalRu_22_D7.g_D07_07.ab1_5	unique	1	35
VspE_030511	16	1_HalRu_22_E2.g_E02_10.ab1_6	unique	2	44
OspG_030511	6	1_HalRu_25_D12.g_D12_08.ab1_4	unique	2	23
OspN_090811	41	1_HalRu_26_B7.g_B07_03.ab1_2	unique	3	14
OspG_030511	47	1_HalRu_27_A10.g_A10_02.ab1_4	unique	1	14
OspN_090811	19	1_HalRu_27_B7.g_B07_03.ab1_2	unique	1	21
OspG_030511	28	1_HalRu_27_F3.g_F03_11.ab1_2	unique	1	20
OspG_090811	35	1_HalRu_28_D6.g_D06_08.ab1_1	unique	3	12
VspE_030511	13	1_HalRu_30_A6.g_A06_02.ab1_5	unique	1	23
OspG_030511	30	1_HalRu_30_E2.g_E02_10.ab1_4	unique	1	17
OspN_090811	36	1_HalRu_31_E2.g_E02_10.ab1_1	unique	1	19
OspN_090811	32	1_HalRu_32_G12.g_G12_14.ab1_5	unique	2	22
OspG_090811	42	1_HalRu_33_2_2_G10_HalRu_33_2_G10.b_14.ab1_5	unique	2	29
OspN_090811	20	1_HalRu_34_2_2_G11_HalRu_34_2_G11.b_13.ab1_3	unique	3	56
OspN_090811	30	1_HalRu_37_B10_HalRu_37_B10.b_04.ab1_5	unique	1	19
OspG_090811	40	1_HalRu_38_B07_HalRu_38_B7.b_03.ab1_6	unique	2	12
OspG_030511	33	1_HalRu_38_B09_HalRu_38_B9.b_03.ab1_6	unique	12	18
OspN_090811	5	1_HalRu_38_E09_HalRu_38_E9.b_09.ab1_2	unique	2	80

VspE_030511	35	1_HalRu_4_C8.g_C08_06.ab1_6	unique	1	13
VspE_030511	45	1_HalRu_4_D12.g_D12_08.ab1_1	unique	3	10
VspE_030511	39	1_HalRu_4_H10.g_H10_16.ab1_5	unique	3	14
OspN_090811	1	1_HalRu_46_E06_HalRu_46_E6.b_10.ab1_5	unique	1	177
OspN_090811	46	1_HalRu_46_G11_HalRu_46_G11.b_13.ab1_4	unique	1	16
OspG_090811	37	1_HalRu_48_G05_HalRu_48_G5.b_13.ab1_4	unique	1	9
OspG_030511	45	1_HalRu_50_B08_HalRu_50_B8.b_04.ab1_2	unique	1	12
VspE_030511	47	1_HalRu_51_C05_HalRu_51_C5.b_05.ab1_3	unique	1	11
OspN_090811	4	1_HalRu_51_H10_HalRu_51_H10.b_16.ab1_3	unique	3	67
OspN_090811	21	1_HalRu_52_B08_HalRu_52_B8.b_04.ab1_2	unique	2	20
OspG_030511	41	1_HalRu_52_E03_HalRu_52_E3.b_09.ab1_2	unique	2	13
OspG_090811	45	1_HalRu_6_C8.g_C08_06.ab1_2	unique	2	10
OspN_090811	27	1_HalRu_6_F12.g_F12_12.ab1_4	unique	1	20
OspN_090811	35	1_HalRu_7_B3.g_B03_03.ab1_3	unique	2	19
OspG_090811	18	1_HalRu_9_A9.g_A09_01.ab1_1	unique	2	31
OspN_090811	42	1_HalRu_9_B11.g_B11_03.ab1_2	unique	1	39
OspN_090811	39	1_HalRu_9_C11.g_C11_05.ab1_3	unique	1	38
OspG_090811	50	1_HalRu_9_H1.g_H01_15.ab1_5	unique	3	11
OspN_090811	9	Contig116_4	unique	3	31
OspG_090811	32	Contig118_2	unique	4	11
OspN_090811	8	Contig123_2	unique	1	36
VspE_030511	44	Contig127_5	unique	1	10
OspG_030511	50	Contig135_1	unique	3	7
OspN_090811	6	Contig138_2	unique	3	40
OspG_090811	7	Contig175_2	unique	5	18
VspE_030511	6	Contig225_1	unique	2	44
OspN_090811	45	Contig225_6	unique	3	20
OspG_030511	49	Contig249_3	unique	2	17
VspE_030511	50	Contig275_5	unique	1	8
OspN_090811	48	Contig325_6	unique	1	12
OspG_030511	38	Contig390_4	unique	1	12
OspN_090811	10	Contig471_2	unique	3	32
OspG_090811	49	Contig482_4	unique	4	13
OspN_090811	49	Contig488_6	unique	2	14

OspG_030511	7	Contig503_1	unique	2	28
VspE_030511	31	Contig544_5	unique	3	16
VspE_030511	43	Contig552_4	unique	1	9
OspG_030511	18	Contig574_3	unique	5	19
OspG_030511	24	Contig578_1	unique	7	44
OspG_090811	46	Contig579_3	unique	3	22
OspG_030511	23	Contig590_3	unique	3	40
OspN_090811	23	Contig613_4	unique	3	32
VspE_030511	23	Contig631_4	unique	2	17
VspE_030511	3	Contig641_6	unique	2	170
OspG_030511	46	Contig645_3	unique	6	20
VspE_030511	30	Contig70_4	unique	1	15
OspN_090811	11	Contig76_6	unique	1	34
VspE_030511	5	gi|21232251|ref|NP_638168.1|	unique	2	24
OspN_090811	15	gi|46575698|gb|AAH69161.1|	unique	2	18
OspG_030511	40	HalFuVEZP5(Contig401)	unique	3	20
OspN_090811	3	HalRu_1_E1.g_E01.ab1_3	unique	2	95
VspE_030511	36	HalRu_2_C12.g_C12.ab1_1	unique	1	27
OspN_090811	14	HalRu_2_G11.g_G11.ab1_3	unique	3	36
OspN_090811	17	HalRu_2_H10.g_H10.ab1_3	unique	1	63
OspG_090811	17	HalRu_36_Testis_G9.g_G09.ab1_3	unique	3	17
VspE_030511	14	HalRu_41_Testis_A3.g_A03.ab1_2	unique	2	23
OspN_090811	38	HalRu_41_Testis_F8.g_F08.ab1_3	unique	1	20
OspN_090811	29	HalRu_42_Testis_A10.g_A10.ab1_5	unique	1	24
VspE_030511	34	HalRu_42_Testis_D1.g_D01.ab1_3	unique	2	16
OspN_090811	34	HalRu_42_Testis_G12.g_G12.ab1_6	unique	2	22
OspN_090811	24	HalRu_44_D10.g_D10.ab1_6	unique	1	26
OspN_090811	50	HalRu_45_E11.g_E11.ab1_6	unique	1	18
VspE_030511	26	HalRu_45_G12.g_G12.ab1_2	unique	3	20
OspN_090811	18	HalRu_ZPL_4	unique	1	38
OspG_030511	17	HalRuZPP	unique	4	29
VspE_030511	2	HalRuZPS_6	unique	2	170
OspG_030511	39	HalRuZPW	unique	6	20
OspG_030511	42	HalRuZPW_1	unique	6	20